3F4-03

半顔テンプレートの学習による複数顔トラッキング

Multiple Face Tracking with learning Half Face Template

松山 純也　　上原 邦昭
Jun'ya Matsuyama　Kuniaki Uehara

神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University

In this paper, we present an algorithm to detect and track multiple human faces in video clips. Both front shots and side shots of a face from video clips can be detected within this algorithm. This algorithm learns Haar-Like features of human faces with InfoBoost algorithm, using these features to detect faces in video clips and track the detected faces with Kalman Filter. To achieve this, we project these features to a 3D model to create the classifier that can detect side shots of a face. However the calculation needs a huge time. So we project the features in the 3D model to a 2D space with respect to many face orientations before the face tracking process begins.

1. はじめに

近年，動画像における人間の顔の検出，追跡に対する要求が高まってきている．たとえば，動画像の撮影範囲内に誰がいるかを認識するには，まず顔がどこにあるかを検出する必要がある．これは，ロボットが人間とコミュニケーションを取る時には，そこに人間がいること，そして相手が誰であるかということを知るために必要である．また，ビルの監視カメラによる入退出の監視などにも適用が期待されているが，これらのシステムでは検出速度も重要な要素となっている．

近年，Viola らによって AdaBoost[2] を用いた高速かつ高精度な複数顔検出を行なうアルゴリズムが提案されている [1]. このアルゴリズムは，単純かつ高速な分類器に AdaBoost を適用して，高速かつ高精度な分類器を生成する．そして，これらの分類器を対象画像の部分画像に適用し，複数の顔を検出するアルゴリズムである．しかし，このアルゴリズムでは，人間の正面の顔しか検出できない．他方向の顔を検出，追跡できるアルゴリズム [4, 5, 6, 7, 8] もいくつか提案されているが，これらのアルゴリズムでは 1 つの顔しか検出，追跡する事ができず，初期パラメータも手動で設定する必要がある．

本研究では，決定理論に基づいた AdaBoost の代わりに，情報理論に基づいた InfoBoost を用いて精度の改善を計ることを検討する．また，顔全体を学習するのではなく，顔の半分のみを学習して，作成された半顔テンプレートを 3 次元モデルに写像し，さらに 3 次元モデルから全顔テンプレートを作成し，鉛直軸中心で回転した顔についても検出できるようにする．

2. Haar 型特徴量

本研究では，画像を分類するための特徴量として，Haar 型の特徴量を使用する．特徴量の例を図 1 に示す．(a) は元の顔画像である．(b) は，両目を横切るような矩形領域とその下の矩形領域を比較すると，目の領域は暗く，その下の領域が明るいという特徴量を表している．また (c) は，両目周辺の矩形領域はそれぞれ暗く，その間の矩形領域は明るいという特徴量を表している．このような単純な特徴量を導入して，画像内の明暗パターンを高速に計算し，画像を分類する．なお，これらの特徴量は図 1 で示されるプロトタイプを縦横に整数倍したものであり，黒領域と白領域の間の明度差をパラメータとして，訓練集合より閾値を求めるようになっている．各特徴量に閾値以上の明度差があれば，対象事例はその特徴を持っていると判断する．

Viola らは図 1 の (a)〜(d) のプロトタイプを使用していたが，本研究ではさらに (e)〜(f) の 4 種類の特徴量を追加している．たとえば，(e) は口とその両端領域に，(f)，(g) は口や目とその上下領域に，(h) は目や鼻，口とその周り等に適合すると考えられる．これらのように，画像の分類において効果的なプロトタイプを追加すると，学習コストは多少増加してしまうが，分類精度は向上すると考えられる．また，(c)〜(h) のような複合的な特徴プロトタイプを追加すれば，分類速度の向上も計ることができる．たとえば，プロトタイプ (e) について考えると，これは (b) を 2 つつないだものと考えることができる．つまり，いままでは (b) の特徴量を 2 回測定しなければならなかったところを，(e) の特徴量を 1 回測定するだけで良くなり，この部分での計算コストを半分にできる．

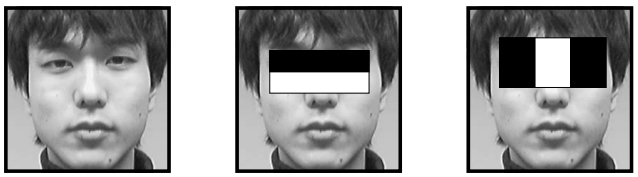

図 1: Feature Example

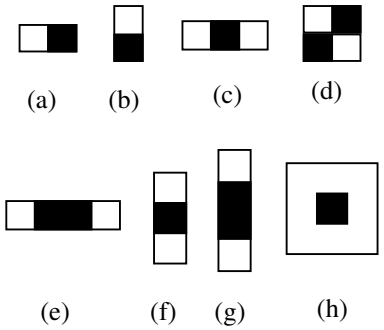

図 2: Feature Prototypes

連絡先: 松山純也，神戸大学大学院 自然科学研究科 情報知能工学専攻，junya@ai.cs.scitec.kobe-u.ac.jp

3. InfoBoost の適用

Haar 型の特徴量による分類は高速であるが，非常に単純であるために精度が低い．Viola らのアルゴリズムでは，AdaBoost によって Haar 型の特徴量による分類を繰り返して，高速かつ精度の高い分類を試みていた．しかし，AdaBoost には 2 つの問題点がある．本研究ではこれらの問題点を解決するために InfoBoost を適用する．

3.1 信頼度

AdaBoost によって生成される各単一分類器は，分類の結果がどの程度信頼できるのかということを考慮していない．そのため，精度の低い分類器も精度の高い分類器も同様に統合されてしまう．たとえば，AdaBoost によって生成された分類器に含まれる各単一分類器が 9 個あり，そのうち 4 個が精度が 95%以上，残りの 5 個が精度が 50%程度だとすると，たとえ 95%以上の精度を持つ分類器 4 個全てが対象画像を「顔」と分類しても，残りの 5 個全てが「顔でない」と分類した場合，対象画像は「顔でない」と分類されてしまう．各分類器の精度を考慮して考えた場合，これは「顔」として分類するべきであるので，ただ多数決を取るのではなく，各単一分類器の精度，信頼度についても考慮すべきである．

3.2 決定理論（decision-theoretic）的

AdaBoost では，重みの更新において，顔を顔として分類した場合と，顔でないものを顔でないと分類した場合を区別しない．つまり，分類が正しかったか，間違っていたかしか考慮していない．しかし実際には，顔を顔として分類した場合，顔を顔でないと分類した場合，顔でないものを顔と認識した場合，顔でないものを顔でないものと認識した場合の 4 つがあり，それぞれの発生確率や分類の正確さは同一であるとは限らない．

たとえば，AdaBoost の途中のラウンドにおいて，顔を顔でないと誤分類した数と，顔でないものを顔と誤分類した数が，それぞれ同程度の数であれば，次のラウンドでの各訓練事例の重みは，分類の正しいものは等しく軽く，誤っているものは等しく重く変更すればよい．しかし，顔でないものを顔と誤分類した数が多く，顔を顔でないと誤分類した数が極端に少ない場合，分類が正しいか否かだけで等しく重みを更新していたのでは，不都合が生じる．顔でないものを顔と誤分類したものについては，以後のラウンドで徐々に正しく分類されるようになっていくと考えられるが，顔を顔でないと誤分類したものについては，誤分類した事例の総数から見た場合の割合が低いため，数ラウンド経っても正しく分類されない可能性がある．これらのことから，分類の正誤だけでなく，実際の分類と分類結果の組み合わせについても考えた上で，重みの更新の際に，どの程度変化させるのかを，4 つの場合それぞれについて考える必要がある．

3.3 InfoBoost の導入

本稿では，上記 2 つの事柄について考慮し，AdaBoost を改良した InfoBoost アルゴリズム [3] を使用する．InfoBoost も AdaBoost と同様に，与えられた WeakLearn を一回呼び出すラウンドを T 回繰り返す（$t = 1, \cdots, T$）が，ラウンド t において WeakLearn によって出力される仮説 h_t は，$h_t : X \to \mathbb{R}$ で示される．これは，この時点で仮説が信頼度付き仮説として出力されているためである．ここでは扱いやすさのため，仮説が取りうる値の範囲を $[-1, +1]$ とする．このとき，h_t の符号は予測するクラス（-1 または $+1$）を表し，絶対値は信頼度の大きさを表す．たとえば $h_1(x_1) = -0.8$ であれば，仮説 h_1 は事例 x_1 を信頼度 0.8 でクラス -1 に分類するという意味である．

次に，この仮説の負予測の正確さ $\alpha_t[-1]$ と正予測の正確さ $\alpha_t[+1]$ を求める．この 2 つの値は，各事例に対する信頼度の大きさや正確さが反映されるようになっている．こうして求めた 2 つのパラメータは，一つのパラメータ α_t として扱われる．$\alpha_t(h_t(x_i))$ は，仮説 h_t が事例 x_i のクラスを -1 と予測したときは $\alpha_t(h_t(x_i)) = \alpha_t[-1]$，$+1$ と予測した時は $\alpha_t(h_t(x_i)) = \alpha_t[+1]$ となる．

分布 D_{t+1} の更新は，$\alpha_t(h_t(x_i))$，ラベル y_i，および信頼度付き仮説 $h_t(x_i)$ を用いて行われる．AdaBoost と同様に，この操作によって正しい分類が容易な事例の重みは低く，困難な事例の重みは高くなるが，信頼度や正予測および負予測の正確さが反映される分，より適切な重みの設定が可能である．

以上の操作を T 回繰り返して，得られた T 個の仮説を用いて最終仮説 H を求めるが，これは信頼度付き仮説 h_t に重みとして $\alpha_t(h_t(x_i))$ を与えたものをすべて足し合わせ，その結果が負ならば -1 を，正ならば $+1$ を最終仮説として出力するものである．

InfoBoost を適用するにあたり，Haar 型の特徴量を用いた各単一分類器は，分類結果がどの程度の正しさをもっているかを示す必要がある．そこで，各訓練事例について特徴量の値を計算し，閾値より大きなもの全体，小さなもの全体それぞれについて実際のクラスの重み付き和（この重みは InfoBoost において用いられている D_t）を求める．その値の絶対値が大きければ，それだけ分類結果は確かなものであり，逆に小さければ，あまり信頼のおけるものではないことになる．すなわち，$+1$ を予測（正予測）された事例，-1 を予測（負予測）された事例，それぞれについて実際のクラスの重み付き和を求めれば，絶対値によって正予測，負予測それぞれが，どの程度信頼できるかが分かる．また，重み付き和の符号は分類結果（-1 または $+1$）の符号に一致する．以上のことから，重み付き和を信頼度付き分類結果として用いている．

InfoBoost をが情報理論に基づいているために，分類精度が上がるという利点の他に，適用対象によって，ある程度の自由度を与えることができるという利点もある．たとえば，監視カメラでの人間の顔検出では，顔を見落とす事は認められないが，逆にある程度の誤検出（顔でない物を顔として検出）は許される．また，ロボットの視覚の 1 機能として使用される場合には，顔を見つけることと顔でないものを顔でないとすることは同程度の重要度であると考えられる．このような各状況ごとの重要性に基づいて，分類の基準を偏らせることができる．

4. 半顔テンプレートの学習と，3 次元モデルへの写像

Viola らのアルゴリズムでは，顔全体を学習して得られた分類器を使用していたため，正面の顔しか検出できなかった．しかし，実際の動画像において，完全に正面を向いた顔の比率はそれほど高くない．そこで，顔の特徴を学習して得られた分類器を，一度，3 次元モデルにマッピングして（図 3(a) → (b)），2 次元的だった分類器を 3 次元的にし，正面だけでなく，横を向いた顔にも対応できるようにする．

しかしながら，3 次元モデルを顔検出に利用すれば，計算量が増大し，速度が大きく低下すると考えられる．そこで，顔検出の前処理として，3 次元モデルを鉛直軸中心に回転させ，分類器を 2 次元に再び写像して，新たな分類器を作成する（図 3(c) → (d)）．これを様々な回転角度において実行しておき，検出段階ではこれらを並列に使用することで計算量の増加を抑える．

また，学習対象として顔全体を使用せず，顔の左右対称性を考え，片側半分のみを学習し，反転させて，学習コストの削減を行なっている（図 3(a)）．半顔の学習と反転によって全顔を表現しているため，光源状態などによる顔の左右の差異についても吸収することができる．

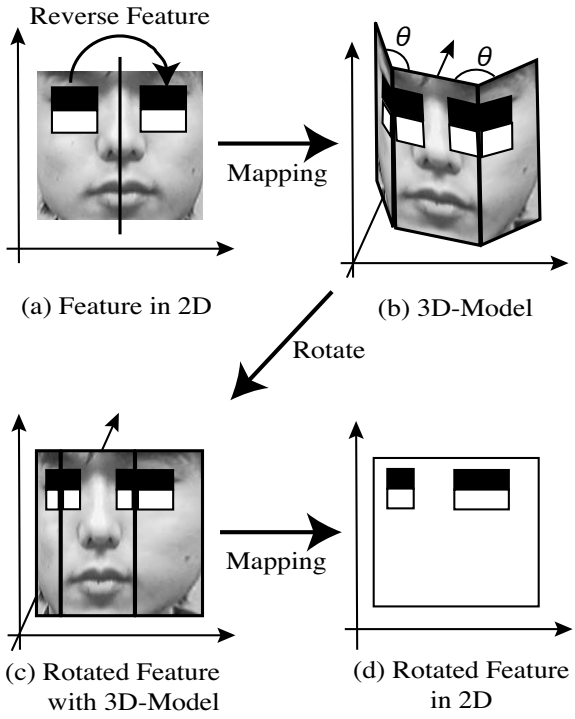

図 3: Feature Mapping

Boosting により得られる分類器は，Haar 型の特徴量を使用する単一分類器の集合であり，各単一分類器は，矩形の基準点の座標と，大きさ，閾値からなる．そのため，図 3(b) で表されるような 3 次元モデルならば，容易にマッピングすることができる．また，3 次元モデルの回転を鉛直軸中心に限定しているので，2 次元に再び写像した際に得られる単一分類器は，元の単一分類器の Haar 型の特徴を横方向に移動，伸縮したものとなる．そのため，Haar 型の特徴量を高速に計算することができる．

また，半顔を用いる際に，Viola らの使用していた Cascading の構成についても変更を行なった．画像の分類は非常に複雑であるため，Boosting によって単一の分類器集合を作成した場合，総数が非常に多くなってしまい，時間がかかる．Viola らは，Cascading によって分類器集合を多段階に分けて分類器の列を作り，いかにも顔らしくないものについては早い段階で除去し，余計な演算を行なわないようにする事で，高速化を計っている．一方，半顔を用いる場合，顔の左右は別々に評価し，評価結果を統合するため，1 つの顔を検出するための分類器が，2 つの分類器列からなる（図 4）．しかし，いずれか一方の分類器列において対象が早い段階で除去される場合，他方の分類器列における評価は無駄になる．そこで，図 5 のように分類器列を交互に行き来する．これにより，2 つの分類器列のうち対象事例の除去が早い分類器列に合わせて処理が終了し，不要な処理が削減される．

3 次元モデルは，実際の人間の顔面形状を反映するために，モーションキャプチャシステムを利用して形状を決定している．通常，モーションキャプチャにおいては人間の関節等にマーカ

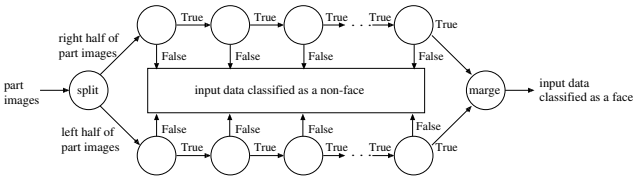

図 4: Two Classifier

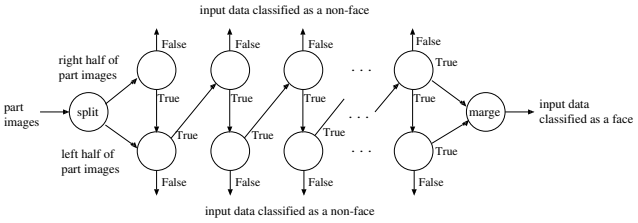

図 5: Join Two Classifier

と呼ばれる反射物を装着し，複数のカメラから測定して 3 次元位置を決定し，人間の動きを導出している．我々は，マーカを人間の顔に貼付け，モーションキャプチャ用カメラで撮影して，顔の 3 次元的形状を測定した．3 次元的形状をそのままモデル化すれば，人間の顔に近いものができあがるが，Haar 型の特徴量の形状が複雑化し，計算量が増大してしまう．そのため，生成される 3 次元モデルは，画像に対して縦，あるいは横方向の直線で折り曲げた形状に限定している．具体的には，人間の口の端から目の中心を通るような直線で約 30 度折れ曲がった図 3(b) のようなモデルを生成している．

5. Kalman Filter の利用

3 次元モデルを使用し，全てのフレームにおいて，全ての部分画像に対して，全ての分類器を適用していたのでは，計算量が膨大になってしまう．そこで，既に検出された顔を追跡する場合には，その前のフレームでの顔の状態パラメータから，現在の状態を類推し，近傍のみを調べて，計算量の増大を防いでいる．具体的には，20 フレームを 1 サイクルとして顔を検出している．各サイクルの 1 フレーム目では全ての分類器を使用している．2〜20 フレーム目では，既に検出された顔の前フレームにおける状態パラメータから，Kalman Filter を利用して現在の状態パラメータを推測する．そして，推測先の近傍に対してのみ顔を検出する．これにより，使用する分類器の数を限定することができ，また，検出対象となる部分画像についても，実際に処理を行なう数を減らすことができる．

6. 実験

現段階では，顔の分類器の 3 次元モデルへの写像，および 3 次元モデルから 2 次元空間への分類器の写像については，実装が完了していない．そこで，既に実装が完了している InfoBoost について，AdaBoost との全顔検出における性能比較実験を行った．

6.1 InfoBoost と AdaBoost の性能比較実験

まず，それぞれの Boosting アルゴリズムを使用した場合の学習プロセスの収束の速さを比較する．結果を図 6 に示す．この図の横軸は「時間（単位は秒）」，縦軸は「顔を顔として検出する割合/顔でないものを顔として検出した割合」である．縦軸の値が小さくなるほど分類器としての性能が高くなる．この図

により，同じ時間の学習を行った場合，InfoBoost は AdaBoost よりも高い性能を示し，また，同程度の性能の分類器を作るためには，InfoBoost のほうが時間がかからないことが示されている．また，学習速度が速いということは，結果として生成される分類器に含まれる単一分類器の数が少ないということである．つまり，InfoBoost は分類速度についても AdaBoost を上回るということになる．

次に，それぞれの Boosting アルゴリズムを使用して生成された分類器の分類精度を比較する．結果を図 7 に示す．この図の横軸は「顔でないものを顔として検出する割合」であり，縦軸は「顔を顔として検出する割合」である．つまり，この図において，グラフが右上にあるほど分類精度が高いということになる．これより，InfoBoost は AdaBoost よりも分類精度が高いということが示されている．以上 2 つの性能比較実験より，AdaBoost の代わりに InfoBoost を利用すれば，学習速度，分類速度，分類精度の全てにおいて性能を向上することが分かる．

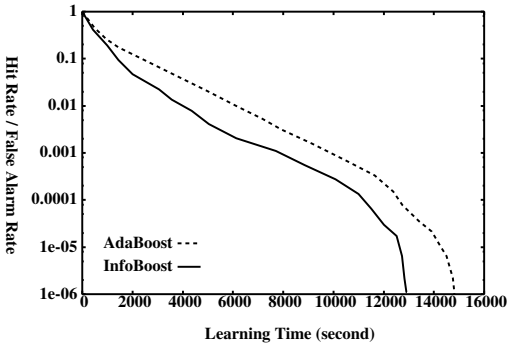

図 6: Learning Speed

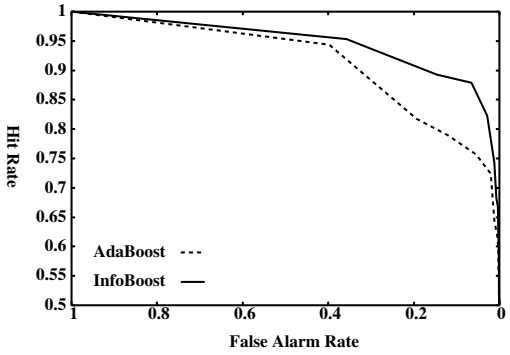

図 7: Classify Rate

7. まとめと今後の課題

本稿では，Viola らのアルゴリズムを元に，InfoBoost を用いて精度の改善を計っている．また，顔の片側半分のみを学習して学習コストを下げている．さらに，作成された半顔テンプレートを 3 次元モデルに写像し，鉛直軸中心で回転した顔についても検出している．しかし現段階では，まだ 3 次元モデル周辺のプログラムの実装が済んでおらず，早急なプログラムの実装が必要である．

また，顔の片側半分のみを学習し，作成した分類器をそのまま使用して，半顔検出を行ったところ，全顔の場合に比べて，検出精度，検出速度ともに大幅に低下してしまった．これは，テンプレートを半顔にしたために，全顔に比べて明度的に特徴的な部分が大幅に減少し，それらを分類するために必要とされる単一分類器が増加し，それでも分類しきれなかったためであると考えられる．訓練画像として 19 × 19 の小さな画像を使用したことも，特徴的な部分の減少，単一分類器の増加を招いた原因と考えられるが，Haar 型特徴量のパターンの追加などによって，より効果的に顔の特徴を捉えることができるようにする必要がある．

また，対象画像からの部分画像の抽出において，肌色領域の検出なども組み合わせれば，さらなる精度改善，速度改善が見込まれると考えられる．また，今回は 20 フレームを単位として，顔の検出と追跡を切り替えるという方法を提案したが，これらを別々のスレッドに分け，共通の顔情報データベースへアクセスして，顔の検出をおこなうフレームで発生する若干の遅延を押さえる事ができると考えられる．その他にも，今回は Kalman Filter を使用したが，それ以外の予測アルゴリズム（Particle Filter）等についても考察する必要があると考えられる．

参考文献

[1] Paul Viola and Michael J. Jones : Rapid Object Detection using a Boosted Cascade of Simple Features, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , pp. 511-518(2001).

[2] Yoav Freund and Robert E. Schapire : Experiments with a New Boosting Algorithm, *Proceedings of the Thirteenth International Conference on Machine Learning (ICML '96)* , pp. 148-156(1996).

[3] Javed A. Aslam, : Improving Algorithms for Boosting, *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory (COLT 2000)* , pp. 200-207(2000).

[4] Zhiwei Zhu and Qiang Ji, : Real Time 3D Face Pose Tracking From an Uncalibrated Camera, *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*(2004).

[5] Ralph Gross, Iain Matthews, and Simon Baker, : Constructing and Fitting Active Appearance Models With Occlusion, *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*(2004).

[6] P. P. Pradeep and P. F. Whelan, : Tracking of Facial Features Using Deformable Triangles, *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision*, Proc. SPIE Vol. 4877, pp. 138-143(2002).

[7] David Ross, Jongwoo Lim, and Ming-Hsuan Yang, : Adaptive Probabilistic Visual Tracking with Incremental Subspace Update, *European Conference on Computer Vision(ECCV 2004)*, pp. 470-482(2004).

[8] Le Lu, Xiang-Tian Dai, and Gregory Hager, : A Particle Filter without Dynamics for Robust 3D Face Tracking, *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*(2004).