

Online Real Boosting による人物追跡

山下隆義^{†1} 勞世竝^{†1} 川出雅人^{†1}

人物の追跡手法として追跡情報の更新を行う際にオンラインで Boosting 学習を行う Online Boosting が注目されている。Online Boosting は弱識別器が多くなると追跡性能が向上するが、処理時間が増大するため、識別能力の高い特徴量を使った弱識別器を選択することが重要である。そこで、我々は識別能力の高い特徴量の導入と効率的に弱識別器を学習するための Boosting を検討した。識別能力の高い特徴量として、Haar-like 特徴量、ABS Haar-like 特徴量、Edgelet 特徴量および EOH 特徴量を組み合わせて用い、学習を行った。また、学習手法として、Adaboost より識別能力の高い弱識別器を学習できる Real Adaboost を用いた。これにより、従来の Online Boosting と比較して、高い追跡性能を得ることができた。また、弱識別器数を 50%削減した場合でも同等の追跡性能を得ることができた。

Human Tracking by Online Real Boosting

TAKAYOSHI YAMASHITA,^{†1} SHIHONG LAO^{†1}
and MASATO KAWADE^{†1}

Online Boosting is an effective incremental learning method which improves the performance of classifier online step by step. It is necessary to use more weak classifiers to construct a strong classifier to attain a high accuracy, however the computation cost also increases. Effective features and weak classifiers selection method should be used to reduce this cost. We use Real Adaboost instead of Adaboost to select features more effectively from a large feature pool consisting of Haar-like feature, ABS Haar-like feature, Edgelet and EOH. We apply this new algorithm to human tracking applications and compare our method with original Online Boosting and Mean Shift. The results show that our method can improve the performance. It is able to keep similar accuracy while the number of weak classifiers is reduced by 50%.

^{†1} オムロン株式会社
OMRON Corporation

1. はじめに

物体追跡はコンピュータビジョンにおける重要なテーマの 1 つであり、特に頭部や人体を追跡対象とした人物追跡の手法が多数提案されている¹⁾⁻³⁾。人物を追跡する方法は大きく 2 つに分けることができる。1 つは、追跡物体の検出処理を毎フレーム行い、検出領域の対応付けから同一物体と見なして追跡する方法である。2 つ目は最初のフレームで物体を検出した後、前フレームでの位置や大きさ、移動量などをもとに時系列変化の情報を用いて追跡する方法である。前者の検出処理として、Boosting を用いて検出器を学習する手法が多く提案され、顔検出において高い有効性が示されている⁴⁾⁻⁶⁾。このような Boosting を用いた手法では、多数の検出対象のサンプル画像を用意し、Haar-like 特徴量を特徴量として用いて学習することにより、照明変動が生じるケースでも検出することができる。また、処理速度が高速であるので、広く用いられている。しかしながら、検出器を用いた人物追跡の場合、顔の向きや傾きが変わると、追跡対象物体の検出ができず、追跡に失敗することが多い。そのため、人物の向きに対応するには向きごとに適した検出器を複数用意する必要がある。検出器の数が多くなると、検出器の数に比例して処理コストが大きくなるため、人物追跡手法は、検出処理を毎フレーム行う手法ではなく、追跡対象に適した特徴量や位置などの時系列情報をもとに追跡する手法が一般的である。このような手法として、Mean Shift²⁾ や Particle Filter³⁾ など多数の手法が提案されており、検出した人物の顔の向きがある程度変化してもそれに合わせて追跡対象の情報を更新することで正しく追跡できる。しかしながら、Mean Shift や Particle Filter による人物追跡手法では、頭部の色ヒストグラムの類似性を用いることが多く、照明変動が生じるシーンや複雑背景下では正しく追跡することができないため、特定の環境に特化して用いることが多い。

近年、これらの問題を解決する物体追跡手法として、物体の検出処理に用いる検出器を学習する Boosting とそれを追跡時にオンラインで更新する検出と追跡を融合した追跡方法が注目されている⁷⁾⁻⁹⁾。Boosting は多数のサンプルを用いて、オフラインで弱識別器を学習する手法であるが、Oza ら⁷⁾ はあらかじめ学習された弱識別器の重みを新しいサンプルデータに合わせて、オンラインで逐次更新していく Online Boosting を提案している。また、Grabner ら⁸⁾ は Online Boosting の枠組みをコンピュータビジョンの領域に拡張して、物体の見え方や隠れが生じるようなシーンにおける物体追跡を実現している。一方で、Avidan⁹⁾ も Oza らと同様に、多数の弱識別器の結果を統合して追跡する Ensemble Tracking を提案している。このように、Boosting を用いたオンライン学習が非常に注目されている。しか

しながら, Online Boosting の計算コストは弱識別器の数に大きく依存しており, 弱識別器を多くすると追跡性能を高くすることができるが, 重みの更新に時間がかかってしまう. 一方, 計算コストを削減するために, 弱識別器数を少なくすると, 性能が低下する. そのため, いかにか識別能力の高い弱識別器を選別し, 少ない弱識別器で十分な性能を達成させるかが重要である. そこで, 我々は複数のタイプの特徴量から識別能力の高い組合せを選択する方法を検討する. また, これらの特徴量の識別能力をさらに高めるために, 従来の Online Boosting で用いられている Adaboost よりも学習能力が高い Real Adaboost¹⁰⁾ を用いた Online Real Boosting について提案する.

本論文では 2 章で Online Boosting について説明し, 3 章で我々の提案する Online Real Boosting について述べる. 4 章では本提案手法を用いた顔追跡と人体追跡の評価結果を示し, 5 章で本手法についてのまとめを行う.

2. Online Boosting

従来の Boosting はあらかじめ用意したサンプルデータをもとに複数の弱識別器を組み合わせた強識別器を学習するオフラインの学習手法の 1 つである. Oza らの Online Boosting は各弱識別器の重みを新たなサンプル画像に適した値に更新していき, 強識別器を最適化していく. Oza らの提案した Online Boosting では, 新しいサンプルに対して, すべての弱識別器の重みを更新するため, 適切でない弱識別器が含まれる場合がある. たとえば, 顔追跡で正面顔に適した強識別器を用意した場合, 顔の向きが横向きに変化すると, 横向きの顔に適していない弱識別器を用いて追跡しようとするため追跡性能が低下する. この問題に対して, Grabner らは, あらかじめ用意された弱識別器をランダムにセレクトとよぶグループに分割し, 各セレクトの中から新たなサンプルに対して最もエラーの小さい弱識別器を識別に適した弱識別器として選び出し, 重みを更新する手法を提案した. 選ばれた弱識別器は新しいサンプルに対するエラーが大きくなると別の弱識別器が選ばれ, 重みだけでなく, 弱識別器もサンプルに合わせて更新される. このように, Grabner らの手法は新しいサンプルにあった特徴量選択も同時に行い, 追跡物体の大きな変化に対する追跡性能を向上させている. しかし, 顔の向きなど物体の形状変化に対する追跡性能を向上させるためには, 多数のセレクトが必要となり, セレクト数が多くなると弱識別器を選択する計算コストが大きくなる. そのため, 少ないセレクト数で識別能力の高い弱識別器を選択することが Online Boosting に必要とされている. そこで, 我々は, このような Online Boosting の課題に対して, 2 つの対策を行い, 少ないセレクト数で高い追跡性能を実現する方法を提案する.

3. 提案手法

3.1 提案手法のフレームワーク

従来の Online Boosting の課題である高い識別能力の弱識別器を選択する方法を, 本論文では弱識別器の特徴量タイプと弱識別器の学習手法の 2 点から検討する. 顔の向きなど物体の形状が変わる場合, 変化によって追跡に有効な特徴量は異なると考えられる. たとえば, 正面を向いている顔を追跡する場合, 顔の濃淡パターンをとらえる Haar-like 特徴量のような特徴量が有効である. 一方で, 横向きなど顔の領域が少ない場合は, 顔の輪郭をとらえるエッジを利用した特徴量が有効である. 提案手法では向き変化に応じて最適な特徴量のタイプをオンラインで選択できるように, 複数の特徴量タイプを用いて, オフラインで学習を行う.

次に, 従来の Online Boosting は AdaBoost により弱識別器を選択している. AdaBoost で選択された弱識別器はしきい値により 0 または 1 の 2 値のみを出力するため, 識別結果の確からしさを得ることができない. そこで, 提案手法では実数による確からしさを出力することで, 高い識別能力を持つ弱識別器を学習できる Real AdaBoost を用いる.

本論文で提案する Real AdaBoost を用いた Online Real Boosting のフレームワークを図 1 に示す. 本手法では追跡したい物体である顔または人体をあらかじめオフラインで学習し, オンライン学習時に選択するための弱識別器群を用意しておく. オフライン学習には Real AdaBoost を用いる. オンライン学習は検出器により検出した位置をもとに学習サンプル

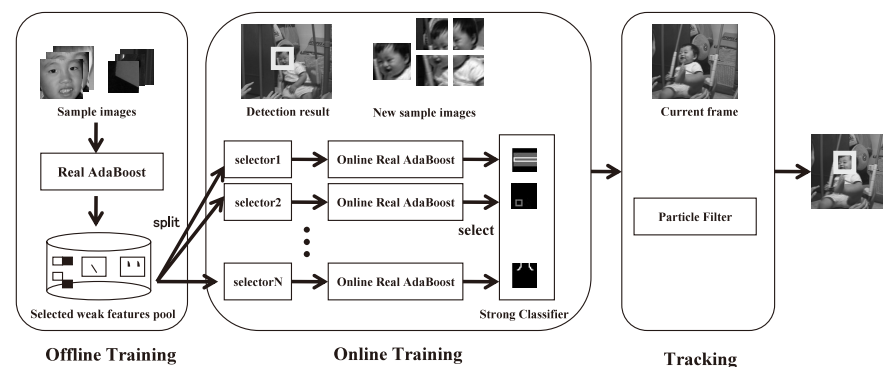


図 1 提案手法のフレームワーク

Fig. 1 Framework of our proposed method.

ルを作成する．次に，あらかじめ学習して得られた弱識別器を複数のセクタに分割し，各セクタの中からサンプル画像を最も正しく判別できる弱識別器を Online Real Boosting により 1 つ選択する．選択された各セクタの弱識別器を組み合わせることで強識別器を作成する．次の時刻に入力された画像からオンラインで学習された強識別器を用い，Particle Filter により追跡物体の位置を検出する．

3.2 特徴量

本節では追跡性能を向上させるために用いた複数の特徴量タイプについて詳しく説明する．また，複数の特徴量タイプを用いた場合の有効性について，顔検出を例に比較した結果を示す．

3.2.1 Haar-like 特徴量

Haar-like 特徴量は Viola ら⁵⁾ の顔検出手法に用いられている特徴量であり，高速に計算できることから，非常に多くの手法で用いられている．Haar-like 特徴量は図 2 に示すように黒の領域と白の領域の 2 つの領域の輝度差を特徴量としている．2 つの領域の取り方により，多くのパターンがあり，図 2 に示すようなパターンが一般的に用いられている．これらのパターンは縦方向または横方向，斜め方向の濃淡変化を特徴量としてとらえている．Haar-like 特徴量の最も優れた特長は積分画像を用いることにより，非常に高速に計算することができることである．特徴量計算のために算出する平均輝度は，黒または白の領域の端点の値を加減算することで，領域内の輝度値の合計を求め，その値を面積で割るだけで算出することができる．

3.2.2 ABS Haar-like 特徴量

ABS Haar-like 特徴量は Haar-like 特徴量と同様に図 2 のようなパターンを用い，2 つの領域の平均輝度の差を特徴量とする．Haar-like 特徴量は 2 つの領域の差が正と負と両方の値を取り得ることがあるが，ABS Haar-like 特徴量は 2 つの領域の差の絶対値を特徴量とする．顔の向きや体の向きなど人物の向きが急激に大きく変化する場合，Haar-like 特徴量は値の正負が反転するが，ABS Haar-like は向きの急激な変化が生じた場合でも，特徴量の符号は変わらない．そのため，ABS Haar-like 特徴量は向き変化に強い特徴量である．

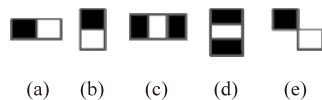


図 2 Haar 特徴量
Fig. 2 Haar-like feature.

3.2.3 Edgelet 特徴量

Edgelet 特徴量は隣接するエッジのつながりを 1 つのパターンとするような特徴量である．Wu ら¹¹⁾ は人体検出における有効な特徴量として Edgelet 特徴量を用いている．Edgelet 特徴量は K 個の画素から構成されており，各要素は位置 u_i とエッジ方向 V_i^E からなる．エッジ方向 V_i^E は図 3 (a) のように 6 方向に量子化された値である．入力画像 I の注目領域 p における Edgelet 特徴量 $S(p)$ は式 (1) に示すように， i 番目の要素のエッジ強度 $E^I(u_i + p)$ と近似類似度関数 $l[\cdot]$ の内積を K 個の要素位置について，累積し正規化した値である．

$$S(p) = \left(\frac{1}{K} \right) \sum_{i=1}^K E^I(u_i + p) l[V^I(u_i + p) - V_i^E] \quad (1)$$

ここで， $u_i + p$ は注目領域 p における Edgelet 特徴量の i 番目の要素位置である．エッジ強度 E^I は，Sobel フィルタを用いて作成したエッジ画像 E の画素値から得ることができる．また，近似類似度関数 $l[\cdot]$ は式 (2) に示すように，入力画像 I の注目領域 p における i 番目の要素位置のエッジ方向 $V^I(u_i + p)$ と Edgelet 特徴量の i 番目の要素のエッジ方向 V_i^E との差をもとに算出される．

$$l[\cdot] = \begin{cases} 1x - = 0 \\ 0.8x = \pm 1, \pm 5 \\ 0.5x = \pm 2, \pm 4 \\ 0x = \pm 3 \end{cases} \quad (2)$$

ここで， $V^I(u_i + p)$ は V_i^E と同様に，6 段階に量子化しており， x はエッジ方向 $V^I(u_i + p)$

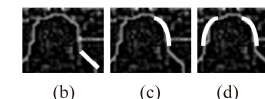
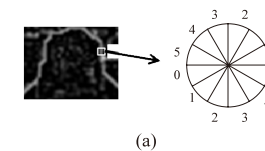


図 3 Edgelet 特徴量
Fig. 3 Edgelet feature.

と V_i^E の差である． x は -5 から 5 までの範囲の値であり，それぞれの値に応じて，式 (2) のような値をとる．

オフライン学習時に選択する Edgelet 特徴量は要素数 K と各要素の位置 u_i を変えることで多数のパターンを生成させることができる．すべての組合せを生成させると膨大なパターンが考えられるため，Wu らは図 3(b), (c), (d) に示すような直線，円弧，対称の 3 種類のパターンを基準とし，生成するパターンに制約を設けている．

3.2.4 EOH 特徴量 (Edge Orientation of Histogram)

Levi ら¹²⁾ は少ない学習データで顔検出に有効な特徴量を得るために，EOH 特徴量を用いた顔検出手法を提案している．EOH 特徴量はまず，図 4 に示すように，エッジ画像を作成し，各画素のエッジ強度と方向を算出する．そして，領域 R における方向 k_1 に対する累積エッジ強度と方向 k_2 に対する累積エッジ強度の累計の比を式 (3) により求め，特徴量 $F(R)$ としている．

$$F(R) = \frac{E_{k_1}(R) + \epsilon}{E_{k_2}(R) + \epsilon} \quad (3)$$

ここで， ϵ は 0 による除算を防ぐための係数である．Levi らはエッジ方向を 4 から 8 方向に正規化することで，効果的な特徴量を得ることができている．特徴量のパターンは比を算出する領域と方向の組合せから多くの特徴量パターンを生成できる．また，Haar-like 特徴量と同様に，エッジ方向ごとにエッジ強度の積分画像を作成すると，累積エッジ強度の計算が 4 点の加減算で行うことができ，高速に特徴量計算を行うことができる．

3.3 特徴量の比較

ここまでで説明した各特徴量の有効性を顔検出の学習を通して比較した．比較方法として，すべての特徴量のタイプを用いた場合を基準とし，各特徴量タイプを除いた場合の検出率の低下を比較した．まず，すべての特徴量タイプを弱識別器の候補として Real Adaboost により 300 個の弱識別器を選び出し，強識別器とする．次に，Haar-like 特徴量を除いた残りの特徴量タイプについて，同様に 300 個の弱識別器を選び出し，強識別器を作成する．以下同様に，ABS Haar-like 特徴量，Edgelet 特徴量，EOH 特徴量をそれぞれ除いた場合の強識別器を作成し，検出率を比較した．評価データとして CMU+MIT データベースを用いた．すべての特徴量による検出率 (All feature) と各特徴量を除いた場合の検出率 (exc. ...) を図 5 に示す．図 5 より EOH を除いた場合，最も検出率が低下していることが分かる．これは頭部や輪郭のエッジが特徴量として非常に有効であるからと考えられる．一方，Edgelet 特徴量は検出率の低下が少なく，他の特徴量タイプに比べて検出性能に対する影響

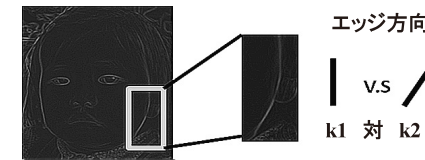


図 4 EOH 特徴量
Fig.4 EOH feature.

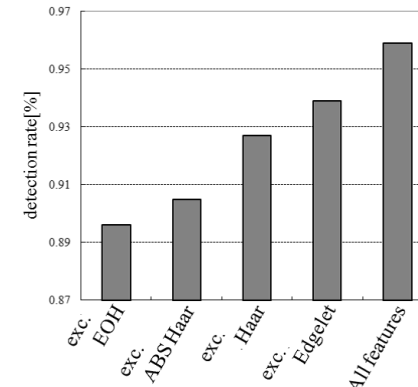


図 5 各特徴量を組み合わせた場合の顔検出の性能
Fig.5 Detection rate of feature set.

度が低い．これは，Edgelet 特徴量は EOH 特徴量に比べて，形状の変化に対するロバスト性が低いと考えられる．Edgelet 特徴量はパターン内の要素ごとにエッジ情報を比較し，判別するのに対して，EOH 特徴量は領域全体のエッジ情報を比較する．そのため，EOH 特徴量はある程度の変化量であれば許容でき，正しく識別できると考えられる．このように，同じエッジ情報を用いた特徴量タイプにもかかわらず，大きな違いが生じている．また，ABS Haar-like 特徴量を除いた場合も検出率が大きく低下している．ABS Haar-like 特徴量は領域間の差の絶対値を特徴量としており，顔の向きが大きく異なる場合でも符号が反転することがないため，正しく識別できていると考えられる．これより，ABS Haar-like 特徴量は向き変化に有効な特徴量であるといえる．

3.4 Online Real Boosting

本節ではここまで説明した特徴量を用いた Online Real Boosting について述べる．提案

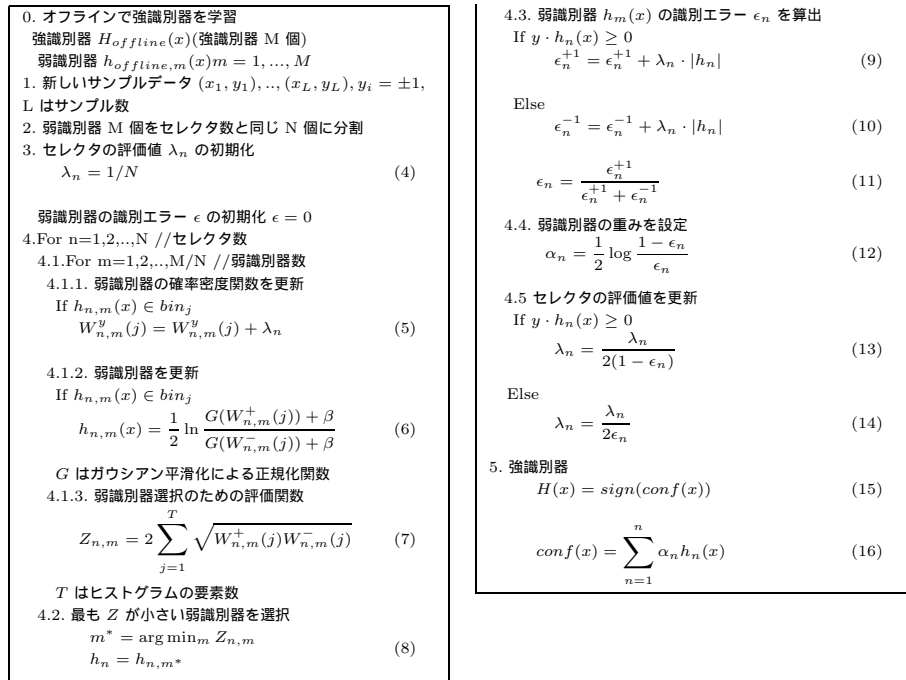


図 6 Online Real Boosting の処理手順
Fig. 6 Procedure of Online Real Boosting.

手法の処理手順を図 6 に示す。Online Real Boosting は、オンライン学習のための準備、弱識別器の選択、弱識別器の更新の 3 つのステップからなる。これらについて詳しく説明する。

3.4.1 オンライン学習のための準備

はじめに、あらかじめ追跡したい物体である頭部の画像を用いてオフライン学習を行い、強識別器 $H_{offline}(x)$ を用意しておく。Grabner らはオフラインで得られた弱識別器セットの中から弱識別器の選択と、学習サンプルの重みの更新をセクタの数だけ繰り返し行う。全セクタが同一の弱識別器セットから選択する場合、サンプルによりセクタ間で同一の弱識別器が選ばれる可能性がある。提案手法は Grabner らと異なり、図 6 の 2 において、弱識別器セットをセクタ数と同じ N 個のグループに分け、各セクタは対応するグルー

プから 1 つの弱識別器を選択する。各グループに含まれる弱識別器セットが異なるようにしており、同一の弱識別器が異なるセクタで選択されることはなく、強識別器の識別能力を高めることができる。また、学習サンプルの重みの更新を行う必要もない。

学習サンプルは図 1 に示すように、正解サンプルに前フレームの頭部領域、非正解サンプルに頭部周囲の背景領域とし、合計 L 個のサンプルを用いる。非正解サンプルの背景領域は頭部位置を一部含むような範囲を設定することで、頭部追跡の追跡位置のずれを軽減することができる。図 6 の 2 において、各セクタの識別能力を示す評価値 λ_n を式 (4) のように初期化する。

3.4.2 弱識別器の選択

次に、図 6 の 4 においてセクタごとに最適な弱識別器を 1 つ選択する。図 6 の 4.1.1 では式 (5) のように、各弱識別器の確率分布 $W_{n,m}^y(j)$ にセクタの評価値 λ_n を重みとして加算して更新する。確率分布 $W_{n,m}^y(j)$ はオフライン学習時に得られた頭部および背景の確率分布であり、正解サンプルは $W_{n,m}^{+1}(j)$ 、非正解サンプルは $W_{n,m}^{-1}(j)$ を更新する。確率分布 $W_{n,m}^y(j)$ はヒストグラムで実現されており、ヒストグラムの要素位置 bin_j は特徴量の値から算出する。全サンプルデータによる確率分布の更新後、式 (6) のように各弱識別器を更新する。弱識別器は確率分布 $W_{n,m}^y(j)$ と同様にヒストグラムで実現されている。サンプルデータにより更新される確率分布の要素が 1 つであると、要素間での確率分布の変動が大きくなる。そのため、1 つのサンプルデータが属する要素は、その前後の要素位置の確率分布を用いて弱識別器を更新するガウシアン平滑化関数を用いる。ここで、 β は 0 による除算を防ぐための係数である。

弱識別器の選択は図 6 の 4.1.3 の式 (7) のようにバチャリア係数 $Z_{n,m}$ に基づく評価関数を用い、4.2 ですべての弱識別器から $Z_{n,m}$ が最も小さな弱識別器をセクタの代表として選ぶ。

3.4.3 弱識別器の更新

選ばれた弱識別器の識別能力を評価するために、図 6 の 4.3 でオンライン学習に用いたサンプルデータにより、エラー率を算出する。正しく識別できた場合、 $y \cdot h_n(x)$ は正の値となり、式 (9) のように正解累積値 ϵ_n^{+1} に加算する。また、負の場合は、誤識別として式 (10) のように非正解累積値 ϵ_n^{-1} に加算する。式 (9)、(10) において、Grabner らは λ_n を加算しているのに対して、本提案手法では $\lambda_n \cdot |h_n(x)|$ を加算している。これは、Real AdaBoost により学習した弱識別器が出力する実数の確からしさをを用いることで、より正確な識別エラーを式 (11) で算出できるためである。式 (11) で得られた識別エラー ϵ_n により、弱識別

器の有効性を示すための値として α_n を式 (12) のように求める．Real AdaBoost は各弱識別器が識別の信頼度を出力するが，本提案手法では，セレクトで選択される弱識別器の有効性を示す α_n を別に算出し，式 (16) において乗じるようにしている．

セレクトの評価値 λ_n は図 6 の 4.5 で識別エラー ϵ_n の値を基に，式 (13) および (14) のように更新する．評価値 λ_n はセレクト n の識別能力を示す値である． λ_n が大きい場合，セレクト n は追跡対象に対して有効な弱識別器を選択しているといえる．一方， λ_n が小さい場合，有効でない弱識別器が選ばれていると考えることができる．本提案手法では，Grabner らと同様に最も λ_n の値が小さいセレクトの弱識別器を更新している．

3.4.4 提案手法と Online Boosting の違い

本提案手法が従来の Online Boosting と大きく異なるのは，各弱識別器が正解および非正解データの確率分布 $W_{n,m}^y(j)$ を持った Real AdaBoost を用いている点である．従来の Online Boosting は AdaBoost を用いているため，各弱識別器が確率分布を持つことができない．これにより，従来手法は新しいサンプルに応じて，弱識別器を選択し，重みを更新するため，追跡物体の形状や照明状態が急激に変化すると，正しく追跡できないことが多い．また，正解・非正解データの確率分布を持たない Real AdaBoost も同様に，新しいサンプルの確率分布を計算し，弱識別器を選択するため，急激な変化が生じた場合は正しく追跡できない．そこで，本提案手法は，あらかじめオフライン学習時に得られた正解および非正解データの確率分布を保持し，新しいサンプルに応じて確率分布を更新している．正解データの確率分布はオフライン学習時に様々な顔の向きや照明変化を学習データに加えることにより，これらを考慮した分布を得ることができる．そのため，頭部の向きや照明状態が大きく変化した場合でも安定して追跡することができる．

3.5 Particle Filter による追跡

本手法での物体検出後の追跡には Particle Filter を用いた³⁾．Particle Filter は時系列フィルタリングの一種であり，対象の状態空間や観測モデルに制約がないため，Particle Filter を用いた多数の追跡手法が提案されている．サンプリング点はそれぞれ重みを持っており，尤度推定により各サンプリング点の重みが更新される．本手法では，Particle Filter により決められたサンプリング点に対して，Online Real Boosting によりオンライン学習した強識別器 $H(x)$ を用いて，各サンプリング点の尤度を計算する．サンプリング点のうち，最も $H(x)$ が大きい位置を人物追跡結果として出力する．Particle Filter の各サンプリング点の重みは，強識別器 $H(x)$ の値をもとに更新する．

4. 実 験

4.1 従来手法との性能比較

本提案手法である Real AdaBoost による Online Real Boosting の有効性を示すために，従来手法である AdaBoost による Online Boosting との頭部追跡性能の比較実験を行った．評価データとして Jepson ら¹³⁾ が評価に使用している動画像 1145 フレームを用いた．まず，頭部検出器をあらかじめオフラインで学習した．学習データは正解データ約 5,000 枚，非正解データ約 9,000 枚を 24×24 ピクセルで正規化し，両手法とも 300 個の弱識別器を学習した．学習データの一例を図 7 に示す．学習データは正面顔から横顔まで様々な向きの顔を含んでおり，頭部の上部と下部の長さが 20 ピクセルになるようにし，背景を一部含むようなサイズで正規化した．提案手法で選ばれた弱識別器の上位 10 個を図 8 に示す．選ばれた弱識別器のうち，(a) と (b) のように白と黒の 2 色で表現されている特徴量は Haar-like 特徴量であり，(c) と (d) のように白線で囲まれた領域で表現している特徴は EOH 特徴量，(f) と (i)，(j) のように黒い領域と白線で表現されている特徴量は ABS Haar-like 特徴量である．

次に，追跡時のセレクト数を 100 とし，あらかじめオフラインで学習した検出器の弱識別器から，オンラインで追跡に有効な弱識別器を選択する．また，Particle Filter のパーティクル数は 200 とし，図 6 の式 (12) で最も信頼度が高い位置を追跡対象領域としている．また，両手法に対して，追跡する頭部の位置は初期フレームで同じ座標を与えている．表 1 に両手法および Mean Shift による追跡対象領域の中心座標の正解座標からの平均誤差と標準偏差を示す．平均誤差は正解頭部の幅に対する正解座標からの位置ずれの割合としている．また，図 9 に各フレームでの正解位置からの誤差と評価データのサンプル画像を示す．表 1

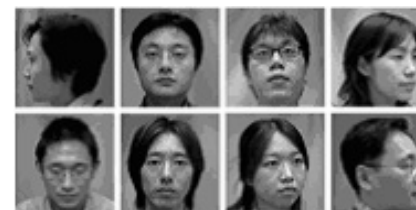


図 7 学習データ例
Fig. 7 Training data samples.

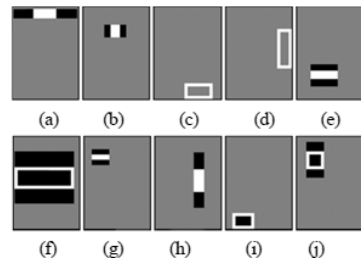


図 8 オフライン学習で選ばれた弱識別機

Fig. 8 Samples of weak classifiers that are selected in offline training.

表 1 追跡性能の比較

Table 1 Performance of each tracking method.

	平均誤差	標準偏差
Mean Shift	11.2%	20.1
Online Boosting	9.8%	6.1
提案手法	6.8%	5.7

より、提案手法が従来手法および Mean Shift に比べて平均誤差が小さく、追跡対象を正しく追跡できていることが分かる。また、提案手法は標準偏差が小さく、追跡対象の動きや隠れ、顔の向き変化が生じた場合でも追跡位置のばらつきが少ない。これらの代表的な結果として、図 9 の 200 フレーム付近および 800 フレーム付近の追跡結果を図 10 に示す。200 フレーム付近では、手による顔の隠れが生じており、Mean shift では手の動きにより追跡位置が大きくずれている。一方、提案手法は追跡位置の誤差が小さいことが分かる。

800 フレーム付近では追跡対象人物が複雑な背景環境へ移動している。そのため、Mean shift では追跡対象と類似する背景に追跡位置がずれている。また、従来手法の Online Boosting においても追跡人物が複雑な背景下へ移動することにより、弱識別器の判別誤りが増加するため、背景の影響を受け、追跡位置が右方向にずれている。一方、提案手法では複雑な背景下でも弱識別器の判別誤りが少ないため、正解に近い位置を正しく追跡できている。これにより、提案手法で用いた複数の特徴量タイプと Real AdaBoost による特徴量選択が有効であることが分かる。

4.2 セレクタ数による性能比較

追跡時に用いるセレクタ数すなわち弱識別器の数による性能を比較する。図 11 にセレク

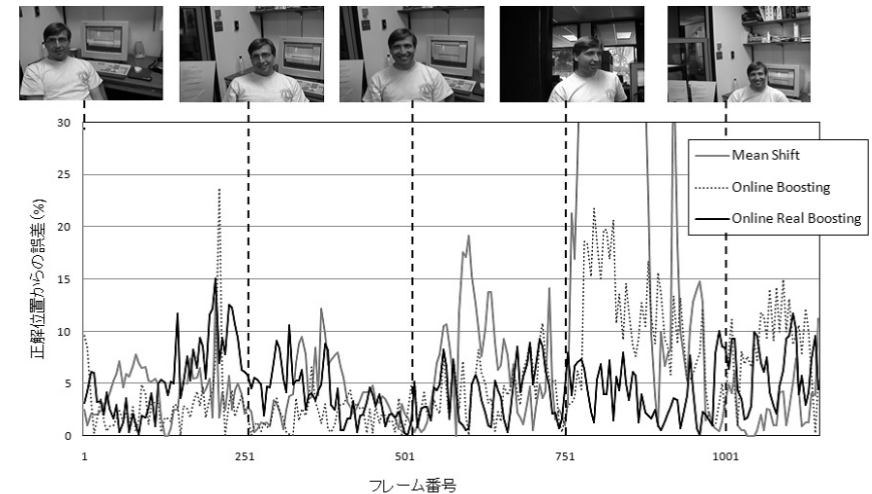
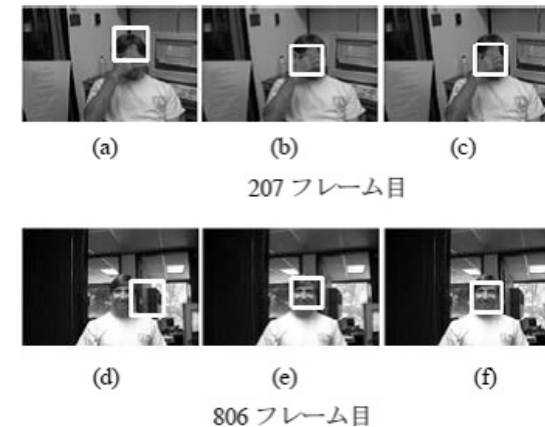


図 9 フレームごとの追跡位置の誤差

Fig. 9 Error of training position in each frame.

図 10 各手法の追跡結果例 (a), (d) は Mean Shift 法, (b), (e) は Online Boosting, (c), (f) は提案手法
Fig. 10 Tracking result of each method.

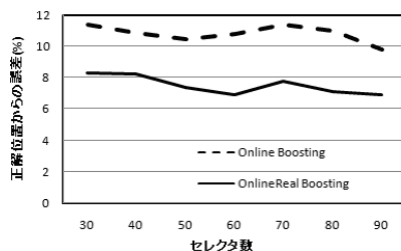


図 11 セレクトタ数による正解位置からの誤差

Fig. 11 The relationship between tracking error and the number of selectors.

タ数を変えた場合の正解位置からの平均誤差の変化を示す．評価データには前節と同様の動画画像を用いた．セレクトタ数を少なくすると，追跡位置の誤差が大きくなる．しかし，提案手法はセレクトタ数を 30 とした場合でも，追跡位置の誤差は 8% 程度であり，Online Boosting でセレクトタ数を 90 とした場合の追跡位置の誤差 9.8% よりも小さい．これより，提案手法は Online Boosting に比べて，少ないセレクトタ数でも安定した追跡が可能であることが分かる．

4.3 顔の向きの変化が生じるシーンでの性能

一般的なデジタルビデオカメラで撮影した動画画像では様々な顔の向きの変化を生じることが多い．特に追跡対象が子供の場合，急激な顔の向きの変化を生じる．本節ではそのようなシーンにおいて，セレクトタ数を少なくした場合の追跡性能を従来手法と比較する．評価データにはデジタルビデオカメラで撮影した動画画像 500 フレームを用いた．表 2 に追跡の中心位置の誤差と標準偏差，図 12 に顔の向きの変化を生じるフレームでの追跡結果の例を示す．図 12 の左側 (a), (d), (g) はセレクトタ数を 100 とした場合の Online Boosting による追跡結果を示している．また，図 12 の中央 (b), (e), (h) は提案手法のセレクトタ数を 50，図 12 の右側 (c), (f), (i) は提案手法のセレクトタ数を 30 とした場合の追跡結果を示している．表 2 より，顔向き変化を生じるシーンにおいても提案手法は，Mean Shift や従来の Online Boosting に比べて，平均誤差が小さいことが分かる．本実験に用いたデータは 46 フレーム付近までは正面を向いており，127 フレーム目付近で斜め左下を向く．以降，265 フレーム目までにかけて，顔を上げて右を向くような動作をする．このような顔の向きが変化するようなシーンにおいて，Online Boosting は十分なセレクトタ数を設定しておくことで，正しく追跡できている．一方，提案手法では，セレクトタ数を従来の Online Boosting より 50% 削減した場合でも同等の追跡性能であることが分かる．これより，提案手法を用

表 2 追跡性能の比較

Table 2 Performance of each tracking method.

	平均誤差	標準偏差
Mean Shift	12.0%	8.6
OnlineBoosting	9.5%	6.3
提案手法 (セレクトタ数 100)	8.2%	5.1
提案手法 (セレクトタ数 50)	8.8%	5.5
提案手法 (セレクトタ数 30)	10.0%	6.4

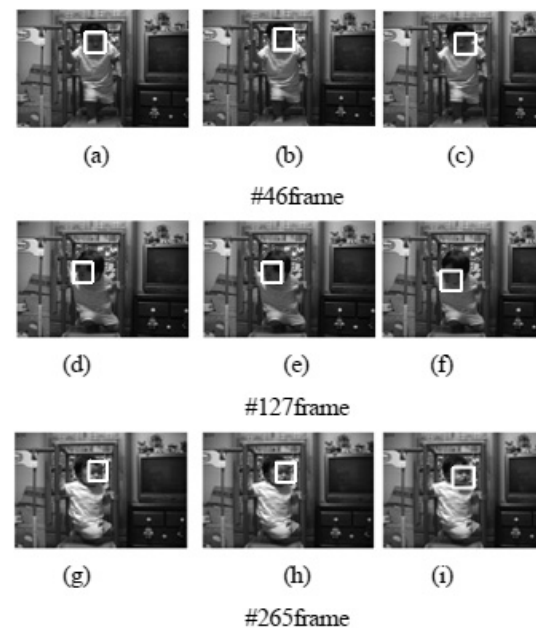


図 12 顔の向きの変化のあるシーンでの性能

Fig. 12 Performance for face pose variation.

いることで，顔の向きの変化が生じるようなシーンにおいても，従来の Online Boosting より少ないセレクトタ数で同等の追跡性能を示すことができた．

4.4 人体追跡への応用

提案手法はあらかじめ追跡したい物体の弱識別器をサンプル画像として学習しておくことで，様々な物体を追跡することができる．本節では，本提案手法を人体追跡に応用した実



図 14 体の向き変化が生じるようなシーンでの人体追跡の結果
Fig. 14 Example of human body tracking with pose variation.

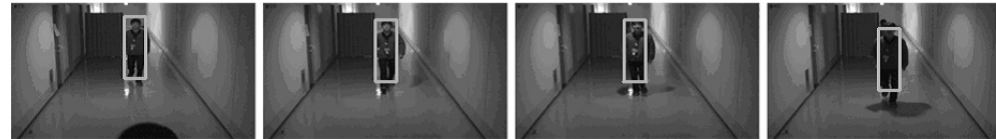


図 15 照明変化が生じるようなシーンでの人体追跡の結果
Fig. 15 Example of human body tracking with illumination variation.

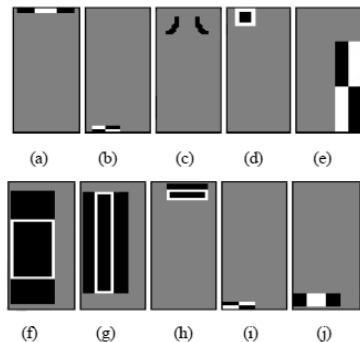


図 13 人体データのオフライン学習で得られた弱識別器

Fig. 13 Samples of weak classifier trained from human training dataset.

験結果を示す．人体データを約 6700 枚で用意し，オフライン学習を行う．人体のオフライン学習は頭部と同様に Real Adaboost により弱識別器 300 個を学習した．図 13 に選ばれた弱識別器のうち上位 10 個を示す．ここで，(c) のように，黒線で円弧のような特徴量が Edgelet 特徴量である．頭部の学習の際は Edgelet 特徴量が上位に選ばれていなかったが，人体の場合は頭部から肩の付近に相当するような位置が選ばれていることが分かる．図 14，

図 15 に提案手法による人体追跡結果を示す．なお，図 14 および図 15 の追跡矩形の色は見やすさの観点から別色としている．図 14 は屋内で体の向きが変化するようなシーンである．体の向きに変化が生じるようなシーンにおいても人体を正しく追跡できていることが分かる．また，図 15 は照明状態が変化するようなシーンである．照明状態が暗く変化した場合でも正しく追跡することができていることが分かる．

5. ま と め

本論文では Online Boosting のフレームワークの中で，識別能力の高い特徴量を選択するために，Real AdaBoost を導入し，Online Real Boosting による物体追跡手法を提案した．また，複数の特徴量タイプを組み合わせることで追跡対象に合わせた有効な特徴量を選択できるようにした．これにより，従来の Online Boosting や Mean Shift に比べて複雑な背景下での追跡性能や顔向き変化が生じるようなシーンで追跡性能が向上していることを確認し，本手法の有効性を示した．また，本提案手法の追跡対象を人体とした場合でも，体の向き変化や照明変化に対するロバスト性を確認することができた．

参 考 文 献

- 1) Li, Y., Ai, H., Yamashita, T., Lao, S. and Kawade, M.: Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Lifespans, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2007).
- 2) Comaniciu, D. and Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis, *IEEE Trans. Pattern Analysis and Image Intelligence*, Vol.25, No.5, pp.603–619 (2002).
- 3) Isard, M. and Blake, A.: Condensation — Conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, Vol.29, No.1, pp.5–28 (1998).
- 4) 勞 世竝, 山下隆義, 岡本卓也, 川出雅人: 高速全方向顔検出, *MIRU2004*, pp.271–276 (2004).
- 5) Viola, P. and Jones, M.: Robust real-time face detection, *International Journal of Computer Vision*, Vol.57, No.2, pp.137–154 (2004).
- 6) Huang, C., Ai, H., Yamashita, T., Lao, S. and Kawade, M.: Incremental Learning of Boosted Face Detector, *Proc. IEEE Conf. on International Conference of Computer Vision* (2007).
- 7) Oza, N. and Rusell, S.: Online bagging and boosting, *Proc. Artificial Intelligence and Statistics*, pp.105–112 (2001).
- 8) Grabner, H. and Bischof, H.: On-line Boosting and Vision, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol.1, pp.260–267 (2006).
- 9) Avidan, S.: Ensemble Tracking, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol.2, pp.494–501 (2005).
- 10) Schapire, R.E. and Singer, Y.: Improved Boosting Algorithms Using Confidence-rated Predictions, *Machine Learning*, No.37, pp.297–336 (1999).
- 11) Wu, B. and Nevatia, R.: Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors, *Proc. IEEE Conf. on International Conference of Computer Vision*, Vol.1 pp.90–97 (2005).
- 12) Levi, K. and Weiss, Y.: Learning Object Detection from a Small Number of Examples: The Importance of Good Features, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.53–60 (2004).
- 13) Jepson, A.D., Fleet, D.J. and El-Maraghi, T.F.: Robust Online Appearance

Models for Visual Tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.10, pp.1296–1311 (2003).

(平成 19 年 9 月 10 日受付)

(平成 20 年 3 月 9 日採録)

(担当編集委員 佐藤 俊雄)



山下 隆義 (正会員)

2002 年奈良先端科学技術大学院大学情報科学研究科修士課程修了。同年オムロン株式会社入社。顔および人画像センシング技術の研究に従事。IEEE-CS 会員。



勞 世竝 (正会員)

1984 年浙江大学電気工学科卒業。1988 年京都大学大学院電気工学研究科修士課程修了。1992 年オムロン株式会社入社。現在、情報処理関連の研究開発、主に顔画像センシング技術に関する研究に従事。IEEE 会員。



川出 雅人 (正会員)

1981 年名古屋工業大学卒業。同年オムロン株式会社入社。現在、顔画像センシング技術に関する研究に従事。