

## Real AdaBoost を利用した Online Boosting による人物追跡

山下隆義、野田隆史、勞世竝、川出雅人  
オムロン株式会社 センシング&コントロール研究所  
[takayosi@ari.ncl.omron.co.jp](mailto:takayosi@ari.ncl.omron.co.jp)

あらまし 物体の追跡手法として物体情報の更新を行う際にオンラインで Boosting 学習を行う Online Boosting 手法が注目されている。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 手法を検討した。識別能力の高い特徴量として、Haar 特徴量、ABS Haar 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 手法を用いた。本手法と Mean Shift を比較した結果、照明変化の大きい状況や向きの変化を伴うシーンにおいて、追跡性能が向上していることを確認することができた。

## Human Tracking by Online Boosting using Real Adaboost

Takayoshi Yamashita, Takashi Noda, Shihong Lao and Masato Kawade  
SENSING&TECHNOLOGY LABORATORY, OMRON Corporation

Online Boosting is an effective incremental learning method which improves the performance of classifier online step by step. It needs many weak learners to construct a strong classifier to attain a high accuracy, however the computation cost also increases rapidly. Effective features and weak learner selection method can 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 Mean Shift. The results show that our method can improve the tracking accuracy especially when there are large illumination changes or pose changes.

### 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、古くから多数の手法が提案されている。物体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、毎フレーム追跡物体を画像全体から検出し、検出領域の対応付けから同一物体とみ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、移動量などの情報をもとに追跡する方法である。前者の場合、物体検出ができると、正しく追従できるが、対象物体の検出が失敗すると追跡できないことが多い。また、毎フレーム探索処理を行うので処理コストも大きい。一方、後者は Mean Shift や Particle Filter など多数の手法が提案されており、検出した物体の形状変化がある程度生じてもそれに合わせて追跡対象の情報を更新すること追跡できる。しかしながら、それだけでは、自動的に追跡を初期化できず、物体を高精度で検出する機能は実アプリにおいて不可欠で、両者を組み合わせることが重要と思われる。本論文では、人の追跡にフォーカスし、顔検出[1][2]と Boosting を使った追跡を組み合わせた人の頭部および人体の追跡手法について検討する。

近年、物体を追跡する手法として、Oza らが提案した Online Boosting が注目されている[3]。Online Boosting は弱識別器の重みを新しいサンプルデータに合わせて、逐次更新していく手法である。また、Grabner らは 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[4]。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従性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。そのため、いかに識別能力の高い弱識別器を選別し、より少ない弱識別器で必要な性能を達成させることが重要である。我々は複数のタイプの特徴量を用いて識別能力の高い組み合わせの探索を行う。また、これらの特徴の識別能力を高めるために、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 Adaboost [5]を用いた Online Real Boosting

について提案する。

本稿では2章で Online Boosting について説明する。次に、3章では本手法で用いた特徴量について説明する。4章で我々の提案する Online Real Boosting 手法を述べる。5章では本提案手法を用いて、頭部追跡と人体追跡の評価結果を示し、7章で本手法についてのまとめを行う。

## 2. Online Boosting

従来の Boosting 手法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。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手法である。しかしながら、Oza らの手法の場合、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従性能が低下する。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する。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。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選択も同時に行い、物体の大きな変化に対する追従性を向上させている。

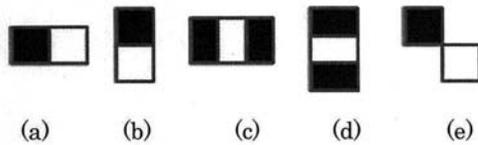


図1. Haar 特徴

### 3. 特徴タイプ

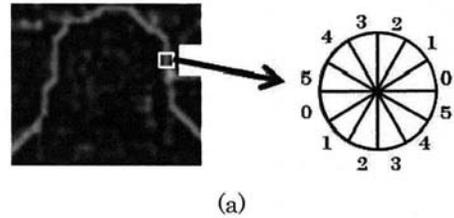
本章では我々が頭部の追従性能を向上させるのに用いた特徴量について下記で詳しく説明する。また、各特徴量を用いた場合の有効性について、顔検出を例に比較する。

#### 3.1. Haar 特徴量

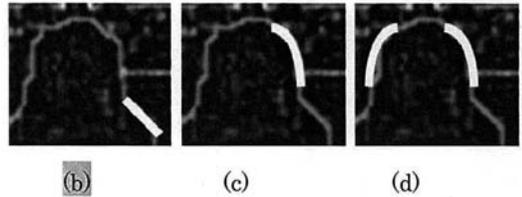
Haar 特徴量はViolaら[6]の顔検出手法の特徴量として用いられている特徴量であり、高速に計算できることから、非常に多くの手法で用いられている。Haar 特徴量は図1に示すように黒の領域と白の領域の2つの領域の平均輝度差を特徴量とする。2つの領域の取り方により、複数のパターンがあり、図1には一般的に用いられるパターンを示す。これらのパターンは縦方向または横方向、斜め方向のエッジに反応するような特徴量である。Haar 特徴量の最も優れた特長は積分画像を用いることにより、非常に高速に計算することができることである。特徴計算のために算出する平均輝度は、黒または白の領域の4点の値を加減算することで、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. ABS Haar 特徴量

ABS Haar 特徴量は Haar 特徴量と同様に図1のようなパターンであり、2つの領域の平均輝度の差を特徴量とする。Haar 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 特徴量は2つの領域の差



(a)



(b)

(c)

(d)

(a) Edgelet 特徴計算のためのエッジ画像とある1点での法線ベクトルの量子化  
 (b)エッジのつながりを直線で表現、(c)エッジのつながりを曲線で表現、(d)エッジのつながりを対照的な線または曲線で表現

図2. Edgelet 特徴

の絶対値を特徴量とする。特長は Haar 特徴量と変わらないが、学習時の特徴量選択において、高速に収束するのに役立つ特徴である。

#### 3.3. Edgelet 特徴量

Edgelet 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[7]ら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。図2に Edgelet 特徴量のパターンの例を示す。特徴はエッジ画像から得た値と法線ベクトルからなる。法線ベクトルは6段階に量子化されている。特徴のタイプは直線、曲線、対称の3パターンあり、つながるエッジの長さや位置により多数の派生したパターンからなる。曲線のタイプは1/2の円弧と1/4の円弧を基本的なパターンとし、直線と同様にエッジの長さや位置から様々なパターンを派生させている。

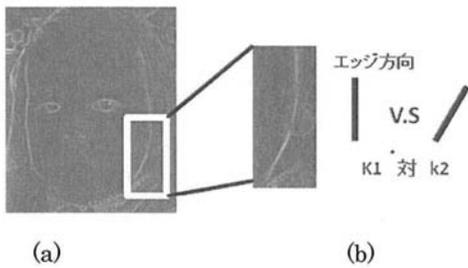


図 3. EOH 特徴

### 3.4. EOH 特徴量 (Edge Orientation Histogram)

Levi ら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[8]。図 3 に示すように、EOH 特徴量はまず、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特徴量とする。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派生させることができる。また、Haar 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。

### 3.5. 特徴量の比較

ここまでで説明した各特徴量が有効であるか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べた。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

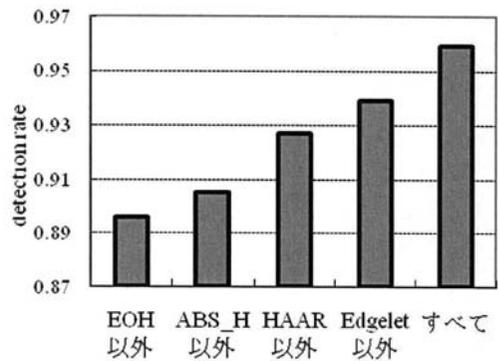


図 4. 各特徴量を組み合わせた場合の顔検出の性能

とする。次に、Haar 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 特徴量、Edgelet 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。評価結果を図 4 に示す。図 4 より EOH を除いた場合、最も検出率が低下していることがわかる。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 特徴量は他の特徴量タイプに比べて検出性能に対する影響度が低い。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ことが考えられる。

Edgelet 特徴はパターン内の各点の法線ベクトル 1 つ 1 つが特徴量であり、各点の比較により判別するのに対して、EOH 特徴は領域全体の法線ベクトルの和を比較する。そのため、EOH 特徴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タイプに関わらず、大きな違いが生じている。しかしながら、Edgelet 特徴は EOH 特徴と同じエッジ画像と法線マップを用いることができるので、無駄な処

理が増加しない。そのため、本手法では上記で説明したすべての特徴タイプを用いることとする。

#### 4. 提案手法

##### 4.1. Online Real Boosting

本章ではここまで説明した特徴量を用いて、効率的に物体を追跡するための Online Real Boosting について述べる。まず、あらかじめ頭部領域についてオフライン学習を行い、強識別器を用意しておく。オンライン学習の開始時に強識別器に含まれる弱識別器を複数のセクタに分割する。次に、前フレーム検出した頭部領域を新しいサンプルデータとし、各弱識別器を更新していく。ここで、各弱識別器の入力は特徴量の値で、出力は追跡対象の確からしさを表す実数である。オンラインで弱識別器の出力関数  $h_{n,m}(x)$  (実装ではルックアップテーブル LUT で実現) を更新していくために、正解データと非正解データの確率分布  $F^{+1}$  と  $F^{-1}$  を保持している。新しいサンプルデータを正解データ、その周囲の領域を非正解データとして与えて、弱識別器  $h_{n,m}(x)$  を更新する。次に新しいサンプルが正しく判別できるか累積エラーを算出する。セクタ内の弱識別器  $m$  個すべてに対して、同様にエラーを算出していき、最もエラーの小さい弱判別器のセクタを代表する弱識別器として用いる。各セクタで選ばれた弱識別器のエラーをそれぞれ算出し、各セクタの重み  $\alpha$  を更新していく。従来の Online Boosting と大きく異なるのは各弱識別器があらかじめ LUT と正解・非正解データの確率分布を持っており、それらの情報をもとに弱識別器を選ぶ点である。これにより、頭部の向きや照明状態が大きく変化しても確率分布は変化していない状態も保持

表 1. Real Online Boosting の処理手順

<p><b>オフライン学習:</b></p> <p>強識別器 <math>F_1</math> (弱識別器 <math>M</math> 個)</p> <p><b>オンライン学習</b></p> <ul style="list-style-type: none"> <li>● 新しいサンプルデータ <math>\mathbf{I}(x, y)</math>, <math>x = 1..N, y \in \{+1, -1\}</math></li> <li>● 重みの初期化</li> <li>● <math>n = 1, 2, \dots, N</math> //セクタ数だけ1から3を繰り返す             <ol style="list-style-type: none"> <li>1. For <math>m = 1, 2, \dots, M</math> //弱識別数                 <ol style="list-style-type: none"> <li>a. 弱識別器の更新 If <math>F_1^m(x) \in \text{bin}_j</math> <math>w_{n,m}^j(i) = w_{1,m}^j(i) + \lambda</math> (2)</li> <li><math display="block">h_{n,m}(x) = \frac{1}{2} \ln \frac{F(w_{n,m}^j(i)+\lambda)}{F(w_{n,m}^j(i)-\lambda)}</math> (3) <math>F</math> はガウシアン平滑化による正規化関数</li> <li>b. 累積エラーを算出 <math>\epsilon_{n,m}</math> <math>y \cdot h_{n,m}(x) \geq 0</math> ならば <math display="block">\lambda_{n,m}^+ = \lambda_{n,m}^+ + \lambda \cdot  h_{n,m} </math> (4) <math>&lt; 0</math> ならば <math display="block">\lambda_{n,m}^- = \lambda_{n,m}^- + \lambda \cdot  h_{n,m} </math> (5)</li> <li><math display="block">\epsilon_{n,m} = \frac{\lambda_{n,m}^+}{\lambda_{n,m}^+ + \lambda_{n,m}^-}</math> (6)</li> </ol> </li> <li>2. 最もエラーの小さい弱識別器を選択</li> <li>3. 重みの更新 <math>y \cdot h_{n,m}(x) \geq 0</math> ならば <math display="block">\lambda = \frac{\lambda}{2(1 - \epsilon_{n,m})}</math> (7) <math>&lt; 0</math> ならば <math display="block">\lambda = \frac{\lambda}{2(\epsilon_{n,m})}</math> (8)</li> <li><math display="block">\alpha = \log_2 \frac{1 - \epsilon_{n,m}}{\epsilon_{n,m}}</math> (9)</li> </ol> </li> <li>● 強識別器: <math display="block">H(x) = \text{sign}(\text{conf}(x))</math> (10) <math display="block">\text{Conf}(x) = \sum_{n=1}^N \alpha_{nh_n(x)}</math> (11)</li> </ul>
--

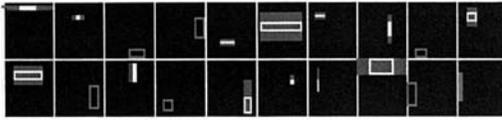


図 5. オフライン学習で選ばれた弱識別器  
300 個の弱識別器を学習したうちの最初に  
選ばれた 20 個

しており、安定している状態と変化の大きい  
状態の両方に対応することができる。

#### 4.2. 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[9]。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、注目されて  
いる手法である。Particle Filter を用いた手  
法が多数提案されている。Particle Filter は  
事後分布 $P(x_t|Y_t)$ と事前分布 $P(x_t|Y_{t-1})$ をそれ  
ぞれのパーティクルの集合 $S_t|S_{t-1}$ 、 $S_t|S_t$ により  
近似する。各サンプルはそれぞれ重みを持って  
おり、尤度推定により重みが決められる。

### 5. 追跡性能

本提案手法による頭部追跡性能について実  
験を行なった。まず、あらかじめ頭部をオフラ  
インで学習を行う。学習は Real AdaBoost によ  
り、正解データ約 5000 枚、非正解データ約 9000  
枚を用意し、300 個の弱識別器を学習した。図  
5 に選ばれた弱識別器の上位 20 個を示す。こ  
こで、最初の 2 つのように白と灰色の 2 色で表  
現されている特徴は Haar 特徴である。3 番目  
と 4 番目に選ばれている灰色の線で囲まれた  
領域で表現している特徴は EOH 特徴である。6  
番目のように灰色と白の線で表現されている

特徴は ABS Haar 特徴である。

次に、追跡時のセクタ数は 50 とし、Online  
Real Boosting 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 100 とし、式(11)の最も信頼度が高い位置  
を追跡対象領域としている。

#### 5.1. 頭部追跡結果

上記で述べたような条件において、頭部追  
跡の性能評価を行った。図 6 に頭部追跡の結果  
を示す。性能を比較するために、Mean Shift  
を用いた場合の結果を(b)に示す。これは、デ  
ジタルビデオカメラで撮影されたホームビデ  
オであり、約 2300 フレームの動画のうち、代  
表的な 4 フレームの結果である。8 フレームで  
は両手法とも正しく追跡手できている。しかし、  
44 フレーム目では顔の向きが下向きになり、  
照明状態が少し暗くなっている。このフレーム  
で、Mean Shift は追跡位置が少し下へずれて  
いるのがわかる。一方で提案手法では正しい位  
置を追跡できている。1 2 3 フレーム目で顔向  
きが横方向へ向くと、提案手法はそれに応じた  
位置を正しく追跡できているが、Mean Shift  
の場合、頭部ではなく、胴体に追跡領域が移動  
しており、追跡が失敗している。さらに、2 5  
8 フレーム目のように、フラッシュが光り顔の  
照明状態が大きく変化すると、Mean Shift で  
は大きく追跡位置がずれてしまう。一方で、  
提案手法では照明変化が生じても追跡位置に影  
響がないことがわかる。このように、提案手法  
を用いることで顔の向きや照明変化が生じた  
場合でも、正しく追跡することができる。

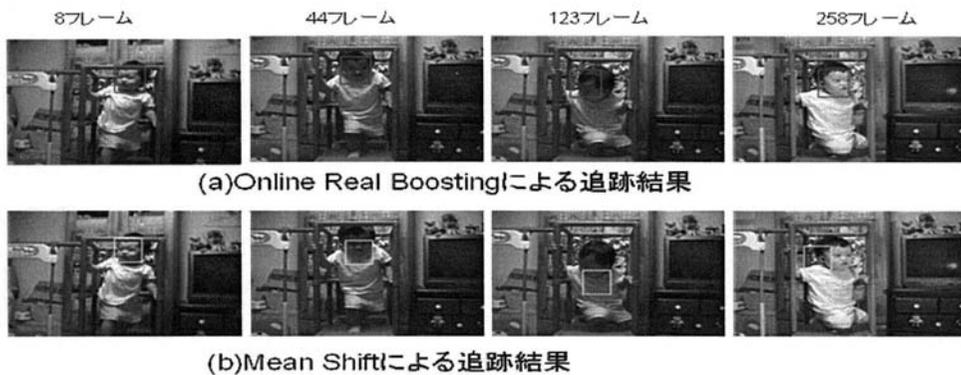


図6. 頭部追跡の結果

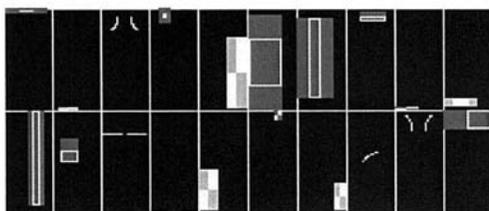


図7. 人体データのオフライン学習で得られた弱識別器

## 6.2. 人体追跡への応用

次に、本提案手法を人体追跡に応用する。人体データを約6700枚で用意し、オフライン学習を行う。サンプル数はある。人体のオフライン学習は頭部の学習と同様に Real Adaboost により弱識別器300個を選びだす。図7に選ばれた弱識別器のうち20個をしめす。ここで、白の直線または円弧のような特徴が Edgelet 特徴である。頭部の学習の際は Edgelet 特徴が上位に選ばれていなかったが、人体の場合は頭部付近や肩の付近に相当するような位置に選ばれていることがわかる。頭部追跡の結果と同様に Mean Shift と比較した結果を図8、図9に示す。図8は体の向きが変化するようなシーンであり、図9は照明状態が変化するようなシーンである。図7より、Mean Shift は体の向

きに変化する364フレーム目において、追跡対象領域がずれてははじめ、412フレーム目では背景の領域を誤って追跡している。一方で、提案手法では体の向きが変化しても正しく人体を追跡することができている。また、図8のようなシーンにおいては、照明状態が大きく変化する40フレーム目より早い29フレーム目で Mean Shift は誤った領域を追跡している。一方で提案手法は照明変化が生じるようなシーンでも正しく追跡できていることがわかる。

## 6. まとめ

本稿では Online Boosting のフレームワークの中で、効率的に特徴量を選択するために、Real AdaBoost を導入し、Online Real Boosting 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、他の追跡手法に比べて照明変化や向きの変化を生じるようなシーンでの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。

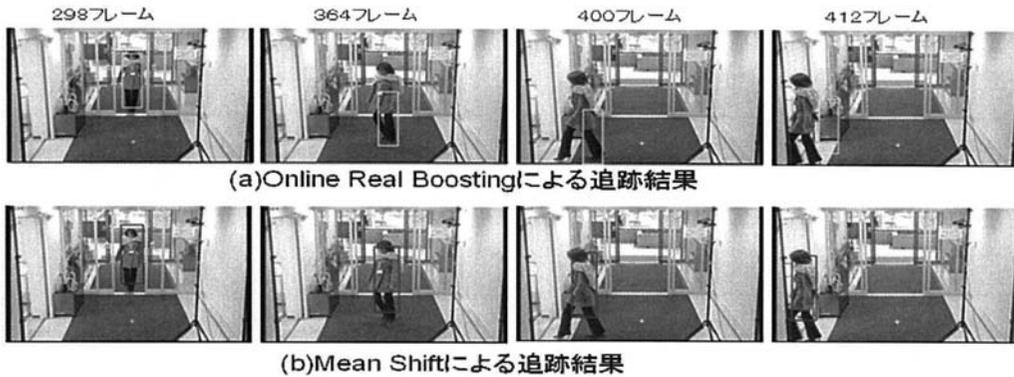


図8. 体の向き変化が生じるようなシーンでの頭部追跡の結果



図9. 照明変化が生じるシーンでの頭部追跡の結果

#### 参考文献

[1] 勞世竝, 山下隆義, 岡本卓也, 川出雅人, “高速全方向顔検出”, MIRU2004, vol. II, pp.271-276, 2004

[2] 山下隆義, 齊藤宏, 平野綾子, 勞世竝, 川出雅人 “組込み環境で動作する高速な顔検出システム”, MIRU2006, デモ発表, 2006

[3] N. Oza and S. Rusell, “Online bagging and boosting”, In *Proc. Artificial Intelligence and Statistics*, pp. 105-112, 2001

[4] H. Grabner and H. Bischof, “On-line Boosting and Vision”, In *Proc. CVPR*, vol. 1, pp. 260-267, 2006

[5] R. E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions”, *Machine Learning*, 37, 1999, 297-336.

[6] P. Viola and M. Jones, “Robust real-time face detection”, In *Proc. ICCV*, vol. .2, pp. 747, 2001

[7] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors”, In *Proc. ICCV*, vol. 1 pp. 90-97, 2005

[8] K. Levi and Y. Weiss, “Learning Object Detection from a Small Number of Examples: The Importance of Good Features”, In *Proc. CVPR*, pp. 53-60, 2004

[9] M. Isard and A. Blake, “Condensation - Conditional Density Propagation for Visual Tracking”, *International Journal of Computer Vision*, vol.29, no.1, pp.5-28, 1998.