

パーティクルフィルタとカスケード型識別器の統合による 人物三次元追跡

小林 貴訓^{†a)} 杉村 大輔[†] 平澤 宏祐^{††} 鈴木 直彦^{††}
鹿毛 裕史^{††} 佐藤 洋一[†] 杉本 晃宏^{†††}

3D People Tracking using the Particle Filter with Cascaded Classifiers

Yoshinori KOBAYASHI^{†a)}, Daisuke SUGIMURA[†], Kousuke HIRASAWA^{††}, Naohiko SUZUKI^{††}, Hiroshi KAGE^{††}, Yoichi SATO[†], and Akihiro SUGIMOTO^{†††}

あらまし 視野を共有する複数のカメラを用いて、三次元空間における人物の実時間追跡を行う。従来より、パーティクルフィルタを用いた人物追跡手法の有効性が報告されている。しかし、観測による仮説の評価は、カラーヒストグラムや輪郭の類似性など、比較的単純な指標が用いられることが多く、実環境での照明変動や複雑背景に対する精度や頑健さなどの点で、必ずしも十分なものではなかった。これに対して本論文では、Haar-like特徴を用いた AdaBoost 学習によるカスケード型識別器を仮説の評価に応用することで、頑健かつ高精度に人物頭部を追跡する手法を提案する。さらに、人物頭部の各方向に対応した識別器を複数準備し、パーティクルフィルタにより生成される仮説と各カメラの関係に基づいて、識別器を適応的に選択することで、人物頭部の向きに伴う見えの変動に対応し、追跡と同時に人物頭部の向きを推定する。実環境における実験により本手法の頑健性、有効性を確認した。

キーワード 人物追跡, 複数カメラ, パーティクルフィルタ, AdaBoost

1. ま え が き

空間内における人物追跡技術は、実世界事象のコンピュータ理解に向けた重要な課題の一つである。特に、カメラからの入力画像を用いて人物を追跡する技術は、監視カメラの普及などに伴い、セキュリティやマーケティングなどへの応用が期待されている。

カメラ画像を用いた人物追跡では、人物が必ずしもカメラに対して一定方向を向いていないために起こる見えの変化の問題や、広範囲の観察により人物が低解像度で観察される問題、昼夜の照明変動や複雑な背景の問題などに対応する必要がある。

カメラ画像を用いた対象の追跡では、これまでに多くの手法が提案されているが、なかでも近年、パーティクルフィルタの有効性が数多く報告されている ([2], [4] ~ [16], [18], [19] など)。パーティクルフィルタは、状態量と尤度を持つ多数の仮説群により離散的な確率密度として追跡対象を表現し、状態遷移モデルを用いて伝播させることで、動きの変動や観測のノイズに対して頑健な追跡を実現する手法である。特に、観測値が非ガウス型となる場合においても頑健な追跡を実現できるとして注目されている。

パーティクルフィルタによる人物追跡では、仮説の人物らしさをカメラ画像を用いて評価するが、これにはカラーヒストグラムや輪郭の輝度変化の類似性などが用いられることが多い ([2], [4], [5], [7], [9], [11] ~ [14], [16], [18])。しかし、このような比較的単純な評価手法は、照明変動や複雑な背景下における人物の追跡では必ずしも十分ではなく、追跡性能の向上には頑健かつ高精度な評価手法が求められる。

一方で、静止画像などから人物の顔を検出する手法が数多く提案されている。なかでも、AdaBoost 学習

[†] 東京大学生産技術研究所 東京都目黒区駒場 4-6-1
Institute of Industrial Science, The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo

^{††} 三菱電機 (株) 先端技術総合研究所 兵庫県尼崎市塚口本町 8-1-1
Advanced Technology R&D Center, Mitsubishi Electric Co. 8-1-1, Tsukaguchi-honmachi, Amagasaki City, Hyogo

^{†††} 国立情報学研究所 東京都千代田区一ツ橋 2-1-2
National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo

a) E-mail: yosinori@iis.u-tokyo.ac.jp

による識別器を用いた顔検出手法が良く知られており、特に、Haar-like 特徴を用いた識別器による顔検出手法は、照明変動や複雑背景下での観察にも頑健であり、実行速度の速さと検出精度から、広く利用されるようになってきている。

AdaBoost 学習による識別器では多数の弱識別器を線形結合することで高精度な識別器を構成するが、Viola と Jones はこれをカスケード型とし、学習、検出時に用いる Haar-like 特徴を高速に計算する手法 [17] を提案している。このカスケード型識別器は、非検出対象がカスケードの初期に棄却されるため、単一の識別対象に関して、より高速な処理が可能である。

静止画像からの顔検出では、このカスケード型識別器を画像全体を網羅するように識別対象の大きさや位置を変化させて適用するが、時系列画像を用いて対象を追跡する枠組みでは、画面全体を逐次探索することは効率的ではない。そのため、パーティクルフィルタの仮説の評価にカスケード型識別器を応用し、探索範囲を限定することは有効であると考えられる。

このような考えに基づいて、本論文では、視野を共有した複数のカメラを用い、状態量として三次元位置と向きを持つ人物頭部を三次元空間内で追跡する手法を提案する。具体的には、1) パーティクルフィルタを用いた追跡の枠組みにおいて、仮説の評価に Haar-like 特徴を用いた AdaBoost 学習によるカスケード型識別器 (以下カスケード型識別器と呼ぶ) を応用する。2) 人物頭部は実際の向きやカメラとの位置関係により、その見えが変化するため、人物頭部の方向毎に学習させた複数のカスケード型識別器を準備し、さまざまな方向で観察される人物頭部に対し、識別器を適応的に選択して評価する。

識別器を用いた追跡の枠組みでは、Support Vector Tracking (SVT) [1] が良く知られている。SVT では画面上で識別器の評価値を最大にする探索を行うが、姿勢変動を伴う三次元物体追跡では、姿勢による見えの変化を 1 つの画面上での評価値の勾配によって追跡することは難しい。また、パーティクルフィルタを用いて対象の見えを詳細に検討することで仮説の評価を行う例として、疎テンプレートマッチングを用いて仮説を評価する手法 [6] があるが、これは対象の解像度が低い場合に適した方法ではない。

カスケード型識別器をパーティクルフィルタと併用した例では、Okuma らの手法 [10] が良く知られている。Okuma らは追跡対象の検出にカスケード型識別

器を用い、仮説の生成に検出結果を利用している。しかし、カスケード型識別器を画面全体に逐次適用するため、リアルタイムでの追跡には至っていない。

Yang ら [19] は、Coarse-to-Fine により段階的に行う仮説の評価の枝刈りにカスケード型識別器を用いているが、2 値分類器としての補助的な利用に留まっている。また、本研究と同時期に、Thierry らは仮説の評価に AdaBoost 学習による非カスケード型識別器を利用する手法 [15] を提案している。しかし、単一の識別器を用いた画面上での追跡に留まっており、姿勢による見えの違いを考慮して三次元空間内で対象を追跡するものではない。

三次元空間内で追跡を行う例では、複数のカメラによる仮説の評価を統合して人物を追跡する手法 [7], [18] や、さらに環境モデルを併用して頑健な追跡を行う手法 [14] が提案されているが、各カメラにおける仮説の評価には、比較的単純な手法が用いられるに留まっている。Nickel ら [8] は、複数のカメラを用いた仮説の評価にカスケード型識別器を利用しているが、カスケード型識別器による検出結果の二次的な利用であり、対象の向きとカメラとの関係は考慮していない。

これに対して本論文では、視野を共有した複数のカメラを用い、三次元空間内で人物頭部を頑健に追跡する手法を提案する。具体的には、パーティクルフィルタの仮説の評価にカスケード型識別器を応用することで、高速かつ高精度に人物頭部を追跡する。また、人物の向きとカメラとの関係に基づいて、識別器を適応的に選択して用いることで、高精度な追跡と同時に人物頭部の向きを推定する。

本稿では、2 節と 3 節にて、パーティクルフィルタとカスケード型識別器について概観し、4 節にて提案手法について詳細に述べる。5 節にて実験と結果について言及し、6 節にて考察を行う。最後に、7 節にてまとめと今後の展望について述べる。

2. パーティクルフィルタ

時刻 t における対象の状態量を \mathbf{x}_t 、画像による観測を \mathbf{z}_t とし、時刻 t までに得られる観測を $\mathbf{Z}_t = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ とする。このとき、時刻 t における対象の事前確率 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$ は、マルコフ過程を仮定することにより、時刻 $t-1$ における事後確率 $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ と時刻 $t-1$ から t への状態遷移確率 $P(\mathbf{x}_t | \mathbf{x}_{t-1})$ を用いて以下のように表すことができる。

$$P(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \int P(\mathbf{x}_t | \mathbf{x}_{t-1})P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})d\mathbf{x}_{t-1}. \quad (1)$$

ここで, $P(\mathbf{z}_t | \mathbf{Z}_{t-1})$ を一定とすると, 時刻 t における事後確率 $P(\mathbf{x}_t | \mathbf{Z}_t)$ は, ベイズの法則より, 時刻 t における尤度 $P(\mathbf{z}_t | \mathbf{x}_t)$ と事前確率 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$ により次式のように表すことができる.

$$P(\mathbf{x}_t | \mathbf{Z}_t) \propto P(\mathbf{z}_t | \mathbf{x}_t)P(\mathbf{x}_t | \mathbf{Z}_{t-1}). \quad (2)$$

対象の追跡は, この事後確率 $P(\mathbf{x}_t | \mathbf{Z}_t)$ の期待値を逐次求めることで実現される.

パーティクルフィルタでは, 時刻 t における事後確率 $P(\mathbf{x}_t | \mathbf{Z}_t)$ を, 状態量 \mathbf{x}_t の仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ と各仮説に対応する重み $\{\pi_t^{(1)}, \dots, \pi_t^{(N)}\}$ により離散的に近似し, 次のプロセスを経て, 逐次的に更新する.

(1) 仮説の選択

時刻 $t-1$ における事後確率 $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ を離散的に近似した N 個の仮説 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ の重み $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ の比に従い, 仮説群 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ を選択する.

(2) 状態遷移確率に基づく伝播

選択された仮説群 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ を, 状態遷移確率 $P(\mathbf{x}_t | \mathbf{x}_{t-1} = s_{t-1}^{(n)})$ に従い伝播し, $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$ に相当する時刻 t における N 個の仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ を生成する.

(3) 画像による重み $\pi_t^{(n)}$ の推定

画像から尤度の評価を行うことで, 仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ の重み $\pi_t^{(n)} \approx P(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)})$ を推定する.

3. カスケード型識別器

Viola と Jones により提案された顔検出器 [17] は, 検出時間の短縮のため, 複数の識別器を組み合わせたカスケード構造 (図 1(a)) をなしている. 図 1(a) において H_i は識別器を表す. 入力画像に対し, 各段で顔, 非顔の判定を行い, 顔と判定された画像だけが次の段へ進む. 最後の段まで通過したものが顔と判定される.

カスケードの各段を構成する識別器 $H_i(x)$ は, 図 1(b) のような特徴を用いて評価を行う多数の弱識別器 $h_t(x)$ の線形結合により, 以下のように表される.

$$H_i(x) = \text{sgn} \left(\sum_{t=1}^T \alpha_t h_t(x) \right). \quad (3)$$

ここで, T は用いられる弱識別器の数であり, α_t は学習

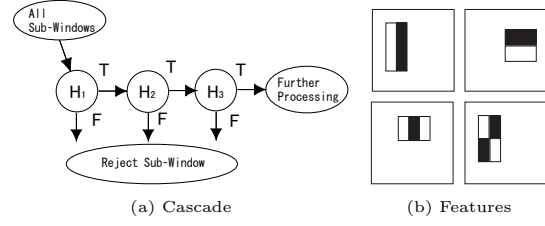


図 1 カスケード型識別器
Fig. 1 Cascaded classifier

時に決まる弱識別器のエラー ϵ_t を用いて $\alpha_t = \log \frac{1-\epsilon_t}{\epsilon_t}$ と表される.

矩形特徴の位置と大きさを画像内でどのようにとるかによって膨大な種類の特徴が存在するが, これらの中から学習用顔画像を良く識別する特徴が AdaBoost アルゴリズムにより選択され, 各段の識別器が準備される.

4. 提案手法

視野を共有した複数のカメラを用いて, 三次元位置と向きを状態量に持つ人物頭部を, パーティクルフィルタにより追跡する. 本論文が新たに提案する内容は以下の通りである.

- 1) カスケード型識別器をパーティクルフィルタの仮説の評価に応用する.
- 2) 頭部の各方向毎に学習を行った識別器を, 仮説とカメラの関係に基づいて適応的に用いることで, 頭部の見えの変化に対応する.

本節では, 提案手法について詳細を述べる.

4.1 人物頭部モデル

室内空間に床面を XY 平面と一致させ, 高さ方向を Z 軸とした三次元世界座標系 XYZ をとる. 人物頭部モデルとして, 中心座標が (x, y, z) である一定の大きさの楕円体を仮定する. 人物は頭部を傾けて室内を移動することは少ないと仮定すると, 人物頭部の向きは Z 軸回りの回転 θ のみで表せる. 仮説は, この四次元の状態量を持つ.

4.2 運動モデルに基づく仮説の伝播

状態遷移確率 $P(\mathbf{x}_t | \mathbf{x}_{t-1})$ として人物頭部の運動モデルを仮定する. 人物頭部の急な動きの変化は少ないと仮定し, 仮説 $s_{t-1}^{(n)} = [x_{t-1}^{(n)}, y_{t-1}^{(n)}, z_{t-1}^{(n)}, \theta_{t-1}^{(n)}]^T$ を $s_t^{(n)}$ に以下のように伝播する.

$$s_t^{(n)} = s_{t-1}^{(n)} + v_t + \omega. \quad (4)$$

ここで、 v_t は過去一定時間の人物頭部の各状態量の平均速度を表す $\dot{x}_t, \dot{y}_t, \dot{z}_t, \dot{\theta}_t$ を要素に持つベクトルである。 ω は平均 0, 共分散行列 Σ_ω を持つガウス雑音であり、 Σ_ω は分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\theta^2$ を対角要素に持つ。

次に、このように伝播した仮説 $s_t^{(n)}$ のカメラ画像への射影を考える。

4.3 カメラ画像への射影

時刻 t の n 番目の仮説 $s_t^{(n)} = [x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)}]^T$ は、校正済みの i 番目のカメラ画像への射影関数を F_i とすると、以下のように射影することができる。

$$\mathbf{p}_{i,t}^{(n)} = F_i \left(\mathbf{s}_t^{(n)} \right). \quad (5)$$

ここで $\mathbf{p}_{i,t}^{(n)}$ は、仮説 $s_t^{(n)}$ の三次元位置を i 番目のカメラ画像へ射影したものである。

このとき、 i 番目のカメラによって観察される相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ は以下のように表される。

$$\theta_{i,t}^{(n)} = \theta_t^{(n)} - \tan^{-1} \left(\frac{[C_i - Ks_t^{(n)}]^y}{[C_i - Ks_t^{(n)}]^x} \right). \quad (6)$$

ここで C_i は i 番目のカメラの XY 座標であり、 K は仮説 $s_t^{(n)}$ から XY 位置成分を取り出す行列である。 $[\]^x$ は X 軸に対応する要素を取り出すことを表す。

最後に、人物頭部モデルの楕円体の幅をカメラ画像へ射影し、 i 番目のカメラで観察される人物頭部の幅 l_i を得る。

このように、人物頭部モデルをカメラ画像へ射影することで、カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ 、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ 、人物頭部の幅 l_i を準備する。次節では、これらを用いた仮説 $s_t^{(n)}$ の評価について述べる。

4.4 カスケード型識別器による仮説の評価

各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の人物頭部らしさをカスケード型識別器を用いて評価する。

カスケードの各段の識別器は、階層が進むにしたがって、より多くの弱識別器 $h_t(x)$ を用いて判定を行う。そのため、より多くの識別器を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ は、より多くの人物頭部の特徴を保持していると考えられる。つまり、カスケードの通過段数と人物頭部らしさには正の相関があると考えられる。そこで、本手法では人物頭部候補領域画像 $g_{i,t}^{(n)}$ をカスケード型識別器に入力した際に通過した識別器の数(カスケード段数)を人物頭部らしさの評価値とする。これは、仮説が実際の人物頭部の状態と大きく離

れて生成された場合、対応する人物頭部候補領域画像 $g_{i,t}^{(n)}$ はカスケードの初期に棄却されるため、計算コストの点からも都合が良い。

一方、各段の識別器の識別性能が一定である場合、識別器が直列に配列されていることから、評価値を通過段数に応じて指数的に増加させることも考えられる。しかし、カスケードの各段に配置される識別器が持つ識別性能は一定ではなく、カスケードの初期を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の多くが、カスケードの最後まで通過する傾向があるため、評価値を指数的に増加させると、カスケードの中盤で棄却された人物頭部候補領域画像 $g_{i,t}^{(n)}$ の評価が追跡に反映されにくい。そのため、本手法では人物頭部候補領域画像 $g_{i,t}^{(n)}$ が通過したカスケード段数により仮説の評価を行う。

具体的には、以下の手順により時刻 t における n 番目の仮説 $s_t^{(n)}$ の評価を行い、重み $\pi_t^{(n)}$ を得る。ただし、カスケード型識別器は、正面、 90° 左向き、 90° 右向きなどの人物頭部の向き毎に、人物頭部と非人物頭部で通過する識別器の数(カスケード段数)に十分な差がつくようあらかじめ学習しておく。

(1) 時刻 t の n 番目の仮説 $s_t^{(n)}$ を i 番目のカメラ画像に射影し、カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ 、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ 、カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を得る。

(2) 仮説 $s_t^{(n)}$ を射影したカメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ を中心に、カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を一辺とする領域を切り出す。ただし、仮説を射影した際にカメラの視野外となる場合、以降(5)までの手順は行わず、重みを一定の小さな値とする。

(3) 切り出した画像のサイズを変更^(注1)し、識別器に入力する人物頭部候補領域画像 $g_{i,t}^{(n)}$ を生成する。

(4) 仮説の射影によって得られた相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ に基づいて、カスケード型識別器を選択する。例えば、正面、 90° 右向き、 90° 左向きの3方向の識別器を用いた場合、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ が $-45^\circ \sim 45^\circ$ の場合は人物頭部正面の識別器が選択され、 $45^\circ \sim 135^\circ$ の場合は 90° 左向きの識別器が選択され、 $-45^\circ \sim -135^\circ$ の場合は 90° 右向きの識別器が選択される。

(5) 人物頭部候補領域画像 $g_{i,t}^{(n)}$ を選択したカスケード型識別器に入力し、人物頭部候補領域画像 $g_{i,t}^{(n)}$

(注1): 各カメラ間の観測情報の精度は一定であると仮定し、評価に用いるカスケード型識別器の識別対象画像サイズを固定(例えば 24×24 ピクセルなど)しているため。

が通過した識別器の数 (カスケード段数) を取得する。ここで得たカスケード段数を対応する仮説の重み $\pi_{i,t}^{(n)}$ とする。例えば、カスケードの最初の段で棄却された場合、重みは 1 となり、全カスケード段数が 40 段で、すべての識別器を通過した場合、重みは 41 となる。

(6) 手順 (1) ~ (5) を各カメラに対して行い、得られた重み $\pi_{i,t}^{(n)}$ を次式により統合する。

$$\pi_t^{(n)} = \prod_i \pi_{i,t}^{(n)}. \quad (7)$$

すべての仮説に対して、式 (7) による重みを計算し、期待値をとることで各時刻の人物頭部の状態量を推定する。

5. 実験

本手法の有効性を確認するため、人物頭部の追跡実験を行った。実験は室内天井に設置した視野を共有する校正済みの IEEE1394 カラーカメラ 2 台 (Point Grey Research 社製 Flea) を用いて行った。各カメラの映像は 640×480 ピクセルで取得し、1 台の汎用 PC (Petium4 3.2GHz, Memory 1GByte) で処理した。

カスケード型識別器として、人物頭部の正面、 90° 右向き、 90° 左向きをそれぞれ検出するように学習を行った 3 種類を用いた。カスケード段数は 40 段とし、識別対象画像サイズは 24×24 ピクセルとした。

運動モデルのパラメータである v_t の要素 $\dot{x}_t, \dot{y}_t, \dot{z}_t, \dot{\theta}_t$ は、過去 10 フレームの追跡履歴から計算し、 Σ_ω の分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\theta^2$ の標準偏差は、それぞれ $\sigma_x = 4\text{cm}, \sigma_y = 4\text{cm}, \sigma_z = 2\text{cm}, \sigma_\theta = 45^\circ$ とした。また、各時刻で生成する仮説数は 200 とした。

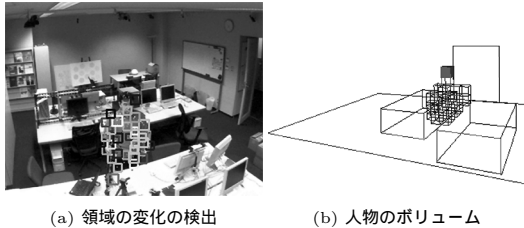


図 2 人物頭部位置の検出
Fig. 2 Detection of head position

追跡は人物頭部候補位置を検出することで開始した。視野を共有するカメラで観測可能な空間を一边 25cm の立方体で分割し、一般的な背景差分により変化が観察された領域のポリウム最上部を人物頭部候補位置として検出した (図 2)。この検出位置に頭部の向きを

一様に分布させた仮説を生成し、追跡を開始した。

追跡の終了は、識別器による評価値の分布が一様に低くなった場合や、推定された人物頭部位置に前景領域が一定以上の割合で検出されない場合とした。

5.1 人物 1 人の追跡

およそ $2\text{m} \times 2\text{m}$ の床上を観測領域とし、人物 1 人の頭部追跡実験を行った。人物は頭部の向きを変えながら、観測領域内を歩き、屈伸動作を行った。この実験では、人物頭部は 1 フレーム (30ms) で平均 1.2cm、最大 6.6cm 移動した。また、カメラ画像では平均 3.3 ピクセル、最大 21.1 ピクセルの変位が観察された。

追跡結果を図 3 示す。頭部位置の推定結果は矩形で画像に重ねて表示した。矩形の色は各カメラからみた人物頭部の向きを示している。図 3 より、ほぼ正確に人物頭部の中心を推定できていることが分かる。また、1 人の人物頭部を追跡した場合、処理は合計 30ms 程度で終了し、リアルタイムでの追跡が可能であった。

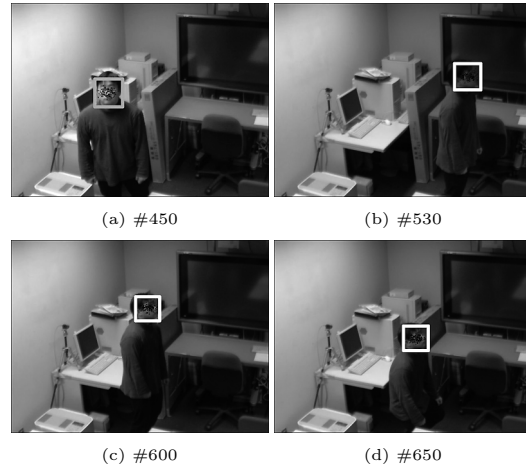


図 3 人物頭部の追跡結果
Fig. 3 Tracking results

提案手法による追跡の精度を定量的に調べるために、画像中の人物頭部位置を手作業で求め、2 枚の画像から逆投影して求めた三次元座標を真の位置と見なし、推定結果と比較した。図 4 に、推定結果と対応する人物頭部の真の位置の XY 平面、及び XZ 平面での軌跡を示す。また、Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 1 に示す。XY 平面上での平均誤差、Z 軸方向の平均誤差は共に 2cm 以内であり、高い精度での追跡を実現している。なお、図 4 において、人物頭部の真の位置の軌跡上に示した灰色の矩形は、その周辺での空間解像度 (画面上の 1 ピクセルの実世界での広がり) を表している。

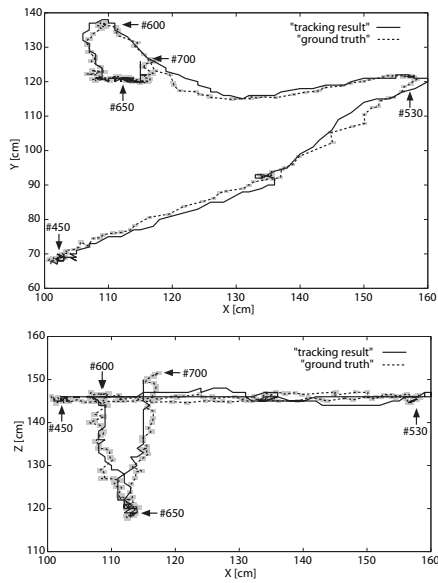


図 4 人物頭部の追跡結果の軌跡
Fig. 4 Trajectory of a user's head position

表 1 人物頭部の追跡における検出誤差
Table 1 Tracking error

	平均 [cm]	標準偏差 [cm]
Z 軸方向	1.02	0.65
XY 平面	1.98	1.46

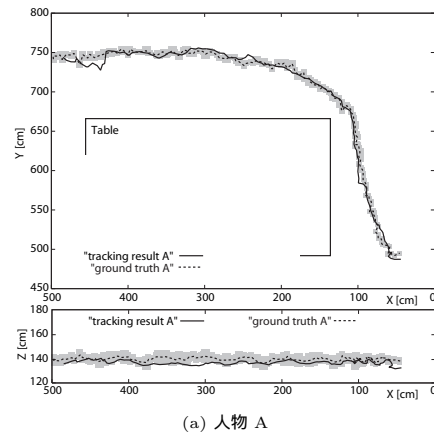
5.2 複数人物の追跡

およそ 5m×5m のより広い床上を観測領域とし、複数人物の追跡実験を行った。3 人の人物は頭部の向きを変えながら観測領域を通過した。この実験では、人物頭部は 1 フレーム (50ms) で平均 5.1cm, 最大 16.5cm 移動した。また、カメラ画像では平均 2.8 ピクセル, 最大 7.1 ピクセルの変位が観察された。

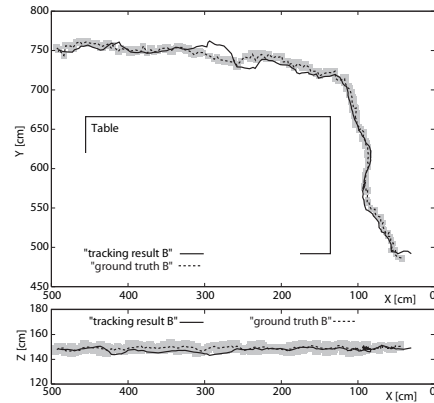
実験の様子を図 5 に示す。図 5 より、複数の人物を、広い観測範囲においても追跡できていることがわかる。また、識別器に入力される画像領域が 20×20 ピクセルと小さい場合でも、追跡を継続することができた。



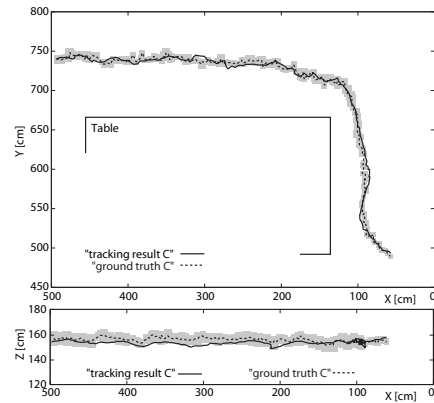
図 5 複数人物の追跡の様子
Fig. 5 Multiple people tracking



(a) 人物 A



(b) 人物 B



(c) 人物 C

図 6 複数人物頭部の追跡結果の軌跡

Fig. 6 Trajectories of users' head position

人物 1 人の頭部追跡実験と同様に、人物頭部位置の推定結果と真の位置の XY 平面及び XZ 平面での軌跡を図 6 に示す。図 4 と同様に、灰色の矩形はその周辺での空間解像度を表している。

また, Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 2 に示す. XY 平面上での平均誤差は 5cm 程度であるが, 観測領域が広く, 画面上での 1 ピクセルが実世界での 5cm 程度に対応する領域が多く含まれることを考慮すると, 高い精度で追跡を実現できているといえる.

表 2 複数人物頭部の追跡における検出誤差
Table 2 Tracking errors

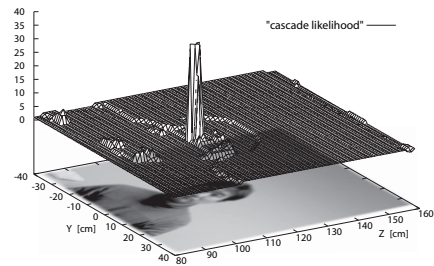
		平均 [cm]	標準偏差 [cm]
人物 A	Z 軸方向	2.59	1.73
	XY 平面	5.08	3.29
人物 B	Z 軸方向	2.14	1.60
	XY 平面	5.63	3.99
人物 C	Z 軸方向	3.77	2.29
	XY 平面	5.03	3.08

複数人物の追跡は Vermaak らの枠組み [16] に基づいている. Vermaak らの枠組みでは, 個々の対象をパーティクルフィルタにより追跡する追跡器を複数動作させ, 追跡器相互の関係に基づいて, 追跡対象のマージなどを行う. しかし, 実験で用いた個々の追跡器における仮説の評価, 及びパラメータは, 人物 1 人の場合と同様である. なお, Vermaak らの手法の詳細は文献 [16] を参照されたい.

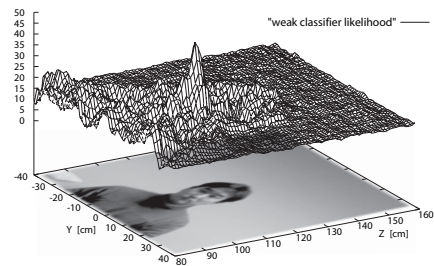
6. 考察

6.1 識別器による仮説の評価

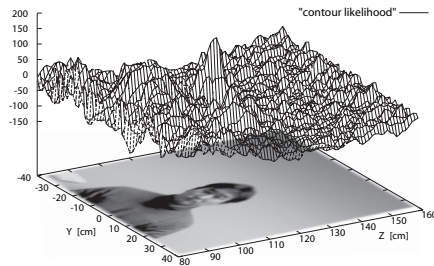
パーティクルフィルタを用いた追跡では, 各フレームにおいて, 仮説の尤度を高精度に推定できることが追跡性能の向上に大きく寄与する. 本手法の枠組みにおいては, 人物頭部の仮説を画像平面上に射影した際に, 人物頭部周辺で鋭いピークを持つ関数が理想的である. そこで, 実際の人物頭部の位置周辺で, カメラから一定距離の平面を一边 1cm のグリッドで分割し, それぞれの三次元位置での人物頭部らしさの評価を提案手法に基づいて算出した. その結果を図 7(a) に示す. また, 比較のために, カスケード段数を 1 段 (カスケード無し) として, 識別器の式 (3) の判定前の値を保存することで人物頭部らしさとした場合の評価 (図 7(b)), 及び, 先行研究で用いられることの多い輪郭の輝度変化の類似性に基づく評価 (図 7(c)) もあわせて示す. なお, 図 7(a) と図 7(b) の識別器は同程度の識別性能を示すように, 同じ学習画像を用いて学習を行った.



(a) カスケード段数に基づく評価



(b) 弱識別器出力に基づく評価



(c) 輪郭の輝度変化の類似性に基づく評価

図 7 評価の分布

Fig. 7 Likelihood distribution

図 7 より, カスケード段数に基づく評価が頭部周辺で高く, 頭部以外の場所では低くなっていることがわかる. また, 輪郭の輝度変化の類似性に基づく評価に比べ, 頭部周辺での評価が鋭いピークを持つことから, パーティクルフィルタでの利用に適しているといえる. また, 図 7(b) の値を人物頭部らしさの評価とすることも十分考えられるが, この方法では, 実際の人物頭部の状態から離れて生成された多くの仮説に対しても, 多数の弱識別器による判定を行うこととなり, 計算コストの面で効率が悪い. しかし, カスケード段数に基づく評価では, 実際の人物頭部の状態から離れて生成された仮説はカスケードの初期で棄却されるため, 計算コストの面で効率が良い.

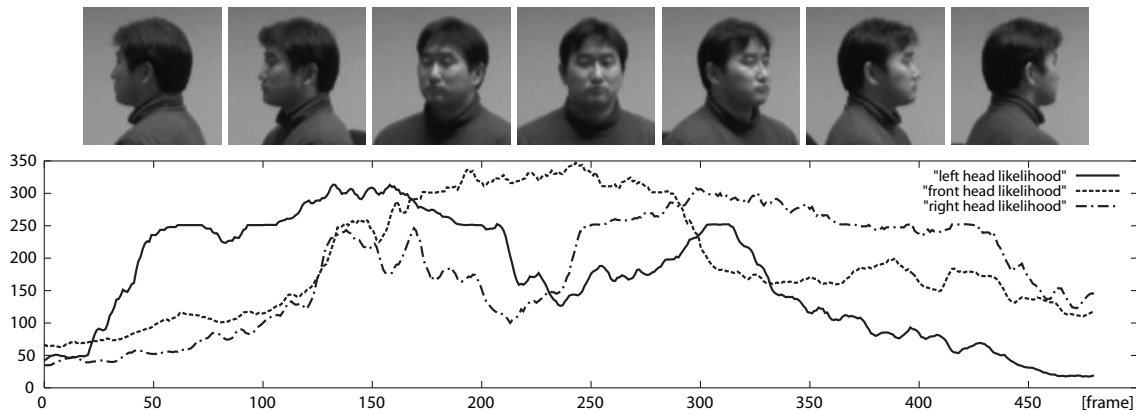


図9 頭部の回転と評価値
Fig.9 Relation of likelihood to head direction

図8は人物1人の頭部を3台のカメラで1200フレームの間追跡した際のカスケードの各段で棄却される仮説の割合を示したものである。各カメラにおいて、カスケードの5段目までに全体の7割以上の仮説が棄却されていることがわかる。

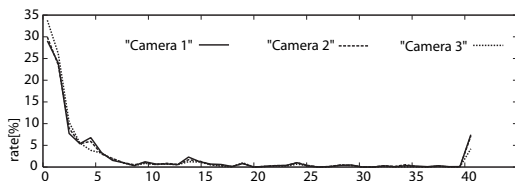


図8 仮説が通過するカスケード段数
Fig.8 Reject samples in each stage

6.2 複数識別器の適応的利用

人物頭部は向きやカメラとの位置関係によりその見えが異なるため、本手法では、複数のカスケード型識別器を仮説とカメラの関係に基づいて選択的に用いている。そこで、相対的な頭部の向きと各方向の識別器の評価の関係を図9に示す。図9は頭部の回転に伴い、評価がどのように変化するかを示したものであり、横軸はフレーム番号、縦軸は頭部周辺を図7と同様に一辺1cmのグリッドで分割した266点の評価値の和を表している。人物はカメラから見て、始めは左を向いており、フレームが進むにしたがって、徐々に正面を向き、最後には右を向いた状態となる。図9上部の写真はグラフの対応するフレーム付近の頭部の画像である。

図9より、頭部がカメラから見て左向きの場合は90°左向き of 識別器の評価が高く、正面、90°右向き

の識別器による評価は低い。頭部が正面を向いている場合は、正面の識別器の評価が最も高く、右を向いている場合は90°右向き of 識別器の評価が最も高い。このように、正しい向き of 識別器の評価が高くなることから、頭部の向き of 仮説に基づいて選択的に識別器を用いても、頭部の向きを正しく推定でき、追跡を継続できる。

複数識別器を適応的に用いることによる頭部の姿勢変動に対する追跡安定性を調べるため、2台のカメラで1人の人物頭部を観察し、人物頭部の正面の識別器1つを用いた場合と、90°右向き、90°左向きを加えた識別器3つを用いた場合で追跡精度を比較した。人物に対して2台のカメラはおおよそ45°の角度を隔てて設置し、人物はどちらか一方のカメラで頭部の正面が観察されるように観測領域内を移動した。図10に1200フレームの間のXY平面上とZ軸方向の追跡誤差を示す。図中の写真は、2台のカメラによるグラフの対応するフレーム付近の頭部の画像である。

400フレームから700フレームの間で識別器1つを用いた場合の誤差が大きくなっている。これは一方のカメラで人物頭部の正面が観察できない状況となり、人物頭部の正面が観察されないカメラにおいて、正しい人物頭部らしさの評価ができていないためと考えられる。これに対して、識別器3つを用いた場合では、頭部の姿勢が変動しても、高精度な追跡が継続できていることがわかる。

本手法により、人物頭部の向きを推定することが可能であるが、正面、90°右向き、90°左向きの3方向の識別器を用いた実験では、正面、右向き、左向きが判別できるに留まり、用いた識別器の数以上の分解能

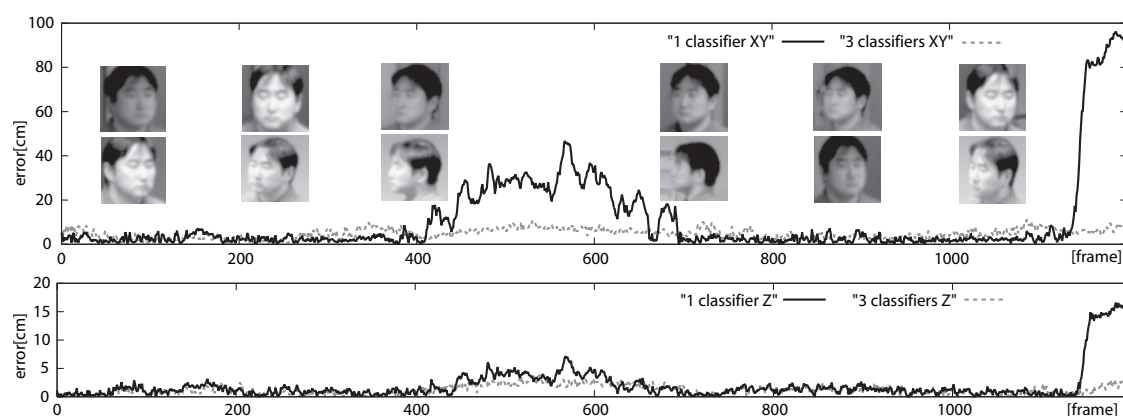


図 10 識別器の数による追跡精度の比較
Fig. 10 Accuracy comparison of single and multiple classifier tracking

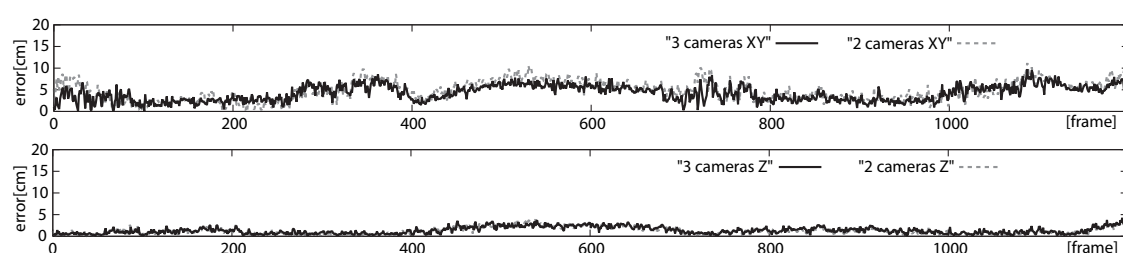


図 11 カメラ台数による追跡精度の比較
Fig. 11 Accuracy comparison of 2 and 3 camera tracking

は観察されなかった．これに対しては，より細かな人物頭部の方向に対応した識別器を追加することで，より細かな頭部の向きを推定できると予想される．なお，本手法ではパーティクルフィルタの仮説により選択的に識別器が用いられるため，識別器の種類を追加した場合でも，計算コストは増加しない．

6.3 カメラの台数と追跡性能

人物を観察するカメラの台数による追跡精度の変化を調べるために，1人の人物頭部の追跡にカメラ2台を用いた場合とカメラ3台を用いた場合で追跡精度を比較した．図11に1200フレームの間のXY平面上とZ軸方向の追跡誤差を示す．

カメラ2台を用いた場合と，カメラ3台を用いた場合では追跡精度に大きな差は観察されなかった．これは，2台のカメラによる仮説の評価により高精度に追跡できているためと考えられる．

また，遮蔽に対する頑健性を調べるために，4台のカメラを用いた追跡実験を行った(図12)．追跡時に1台のカメラにおいて障害物による遮蔽が起きた場合でも追跡は安定していた(図12(a))．これは，一部のカメラで障害物により頭部が遮蔽された場合，遮蔽の起

きたカメラでの識別器による評価が低い値となるが，頭部が観察されるカメラでの評価に，その影響が及びにくいと考えられる．また，本手法では仮説の状態量として三次元位置を持つため，図12(b)のように人物が一部のカメラの視野から外れた場合においても，仮説をカメラに射影することで，視野から外れた仮説を判断することができるため，安定して追跡を継続することができる．

7. むすび

本論文では，視野を共有した複数のカメラを用い，人物頭部を三次元空間内で追跡する手法を提案した．パーティクルフィルタにおける仮説の評価にカスケード型識別器を応用し，複数の識別器を仮説とカメラとの関係に基づいて適応的に用いることで，人物が頭部の向きを変えながら移動する場合でも，高精度に人物頭部を追跡できることを示した．

本手法では，経験に基づいてカスケード型識別器の通過段数から人物頭部らしさを評価しているが，より理論的な裏づけに基づいた識別器による人物頭部らしさの評価について，継続して検討を進めたい．



(a) 障害物による遮蔽



(b) 視野外への移動

図 12 複数カメラによる追跡の頑健性

Fig. 12 Robustness of multi-camera tracking

また、本手法の発展として、同一物体の多視点画像上で得られる観測情報の統合において、観測情報の信頼性を考慮した手法への拡張などが考えられる。さらに、本手法ではあらかじめ学習した識別器を用いて追跡を行っているが、Helmutら [3] により提案されているように、追跡と同時に識別器の学習を行う手法に拡張することについても検討したい。

今後は、より頑健で実用的な人物追跡システムの構築を目指し、店舗などの特定領域内の人物の追跡を想定した複数の観測領域を統合的に扱う枠組みの構築を行う予定である。

文 献

- [1] S. Avidan, "Support Vector Tracking," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.8, pp.1064-1072, 2004.
- [2] S. Birchfield, "Elliptical Head Tracking Using Intensity Gradients and Color Histograms," Proc. the IEEE International Conference on Computer Vision

- and Pattern Recognition, pp.232-237, 1998.
- [3] G. Helmut, G. Michael, B. Horst, "Real-Time Tracking via On-line Boosting," Proc. the British Machine Vision Conference, vol.1, pp.47-56, 2006.
- [4] 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.
- [5] G. Loy, L. Fletcher, N. Apostoloff and A. Zelinsky, "An Adaptive Fusion Architecture for Target Tracking," Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp.261-265, 2002.
- [6] 松原康晴, 尺長健, "疎テンプレートマッチングとその実時間物体追跡への応用," 情報処理学会論文誌, vol.46, no.SIG CVIM 11, pp.60-71, 2005.
- [7] 松本郁佑, 加藤丈和, 和田俊和, "複数カメラを用いた Condensation によるオクルージョンにロバストな人物追跡," 画像の認識・理解シンポジウム (MIRU2006), pp.501-506, 2006.
- [8] K. Nickel, T. Gehrig, R. Stiefelwagen and J. McDonough, "A Joint Particle Filter for Audiovisual Speaker Tracking," Proc. the 7th international conference on Multimodal interfaces, pp.61-68, 2005.
- [9] K. Nummiaro, E. Koller-Meier and L. Van Gool, "An Adaptive Color-Based Particle Filter," Image and Vision Computing, vol.21, no.1, pp.99-110, 2003.
- [10] K. Okuma, A. Taleghani, N. Freitas, J. Little and D. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking," European Conference on Computer Vision, vol.3021 of LNCS, pp.28-39, 2004.
- [11] P. Prez, J. Vermaak and A. Blake, "Data Fusion for Visual Tracking with Particles," Proc. of the IEEE, vol.92, no.3, pp.495-513, 2004.
- [12] J. Sherrah and S. Gong, "Fusion of perceptual cues for robust tracking of head pose and position," Pattern Recognition, vol.34, no.8, 2001.
- [13] 杉本晃宏, 谷内清剛, 松山隆司, "確信度付き仮説群の相互作用に基づく複数対象追跡," 情報処理学会論文誌, vol.43 no.SIG CVIM 4, pp.69-84, 2002.
- [14] 鈴木達也, 岩崎慎介, 小林貴訓, 佐藤洋一, 杉本晃宏, "環境モデルの導入による人物追跡の安定化," 電子情報通信学会論文誌 DII, vol.J88-DII no.8, pp.1592-1600, 2005.
- [15] C. Thierry, V. G. Belille, F. Chausse, J. Thierry, "Real-Time Tracking with Classifiers," International Workshop on Dynamical Vision in conjunction with ECCV, 2006.
- [16] J. Vermaak, A. Doucet, P. Perez, "Maintaining Multi-Modality through Mixture Tracking," Proc. the IEEE International Conference on Computer Vision, vol.2, pp.1110-1116, 2003.
- [17] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.511-518, 2001.

- [18] Y. Wang, J. Wu and A. Kassim, " Particle Filter for Visual Tracking Using Multiple Cameras," Proc. IAPR Conference on Machine Vision Applications, pp.298-301, 2005.
- [19] C. Yang, R. Duraiswami and L. Davis, " Fast Multiple Object Tracking via a Hierarchical Particle Filter," Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.212-219, 2005.

(平成 18 年 10 月 11 日受付, 19 年 2 月 8 日再受付)



小林 貴訓

2000 年電気通信大学大学院情報システム学研究科情報システム運用学専攻修士課程修了。同年, 三菱電機 (株) 設計システム技術センター入社。現在, 東京大学大学院情報理工学系研究科電子情報学専攻博士課程在学中。



杉村 大輔

2007 年東京大学大学院情報理工学系研究科電子情報学専攻修士課程修了。現在, 同大学同研究科同専攻博士課程在学中。コンピュータビジョンに関する研究に従事。



平澤 宏祐

1994 年北海道大学理学部物理学科卒業。1996 年同大学大学院工学研究科修士課程修了。同年, 三菱電機 (株) 入社, 現在に至る。同社先端技術総合研究所にて, 主にヒューマンインタフェース, 生体計測システム, 3 次元計測システムの研究開発などに従事。日本 ME 学会, 日本人間工学会会員。



鈴木 直彦

1997 年東京大学大学院工学系研究科修士課程修了。同年, 三菱電機 (株) 入社。現在に至る。オペレーションズリサーチ, パターン認識に関する研究に従事。現在, 東京大学大学院工学系研究科博士課程在学中。



鹿毛 裕史

1990 年京都大学大学院工学研究科修士課程修了 (情報工学専攻)。同年, 三菱電機 (株) 中央研究所入社。1991-1992 年大阪大学医学部研究生 (神経生理学)。画像認識, ニューロコンピュータ, 人工網膜 LSI, 脳の視覚モデルに関する研究に従事。現在, 三菱電機 (株) 先端技術総合研究所・センサ情報処理システム技術部・画像認識システムグループマネージャ。情報処理学会, 映像情報メディア学会会員。



佐藤 洋一 (正員)

1997 年カーネギーメロン大学計算機科学部ロボティクス学科博士課程修了。Ph.D in Robotics。同年より東京大学生産技術研究所研究機関研究員, 講師, 助教授を経て, 現在同大学大学院情報学環准教授。コンピュータビジョン, ヒューマン・コンピュータ・インタラクション, コンピュータグラフィックスに関する研究に従事。平 18 電子情報通信学会論文賞, 平 11 山下記念研究賞, 平 11 年度日本バーチャルリアリティ学会論文賞等を受賞。情報処理学会, 日本バーチャルリアリティ学会, ACM, IEEE 各会員。



杉本 晃宏 (正員)

1987 年東京大学工学部計数工学科卒業。1989 年同大学院工学系研究科修士課程修了 (数理工学専攻)。日立製作所基礎研究所, ATR, 京都大学を経て, 2002 年より国立情報学研究所。現在, 同研究所教授。総合研究大学院大学複合科学研究科教授併任。博士 (工学)。視覚情報処理や離散システム・アルゴリズムなどに広く興味をもち, 数理的手法に基づいた手法を確立する研究に従事。2001 年情報処理学会論文賞。

Abstract In the last decade, a variety of tracking techniques based on particle filters have been proposed. However, most of the existing methods adopt only simple perceptual cues such as color histograms or contour similarities for hypothesis evaluation. To improve the robustness and accuracy of tracking, it is very important to have a better means for hypothesis evaluation. In this work, we present a novel hypothesis evaluation technique for human head tracking using cascaded classifiers based on AdaBoost and Haar-like features. In particular, we use multiple classifiers, each of which is trained for detecting human heads of a particular direction. Among these classifiers, the most suitable one is selected adaptively by considering each hypothesis and known camera position. Our experimental results demonstrate the effectiveness and robustness of our method.

Key words people tracking, multiple cameras, particle filter, AdaBoost