

行動履歴に基づく人物存在確率の利用による 人物三次元追跡の安定化

杉村大輔^{†1} 小林貴訓^{†1,*1}
佐藤洋一^{†1} 杉本晃宏^{†2}

本稿では、人物の行動履歴を用いた人物追跡の安定化手法を提案する。ある決まった通路の通行、滞留などの人物の行動は、対象空間内の特定の領域で頻繁に観測される。このような人物の行動を長時間観測することにより、行動履歴に基づいた人物の存在確率分布（環境属性と定義する）を得ることができる。そしてこの環境属性を importance function としてパーティクルフィルタの枠組みに組み込むことにより、安定な人物追跡、特に高速な追跡初期化を実現する。また、環境属性は毎フレーム得られる追跡結果を用いて逐次的に更新される。実環境における実験により、本手法の有効性を確認した。

Incorporating Long-term Observations of Human Actions for Stable 3D People Tracking

DAISUKE SUGIMURA,^{†1} YOSHINORI KOBAYASHI,^{†1}
YOICHI SATO^{†1} and AKIHIRO SUGIMOTO^{†2}

We propose a method for enhancing the stability of tracking people by incorporating long-term observations of human actions in a scene. Basic human actions, such as walking or standing still, are frequently observed at particular locations in an observation scene. By observing human actions for a long period of time, we can identify regions that are more likely to be occupied by a person. These regions have a high probability of a person existing compared with others. The key idea of our approach is to incorporate this probability as a bias in generating samples under the framework of a particle filter for tracking people. We call this bias the environmental existence map (EEM). The EEM is iteratively updated at every frame by using the tracking results from our tracker, which leads to more stable tracking of people. Our experimental results demonstrate the effectiveness of our method.

1. はじめに

コンピュータの高速化やカメラの小型化、低価格化にともない、カメラ映像を用いた応用技術が注目されている。なかでも人物追跡技術は、防災・防犯を目的とするセキュリティ分野、効果的な店内レイアウトの設計に向けたマーケティング分野など、様々な分野から大きな期待が寄せられている。しかしながら、実環境においては、障害物や人物相互による遮蔽、複雑背景、見えの変化などの様々な問題があり、いかなる状況においても安定に人物追跡を実現することは難しい。カメラ映像を用いた人物追跡技術はこれまでに多くの手法が提案されているが、なかでもパーティクルフィルタ⁴⁾の有用性が近年数多く報告されている。パーティクルフィルタによる人物追跡は、追跡対象を表す確率密度分布を状態量と尤度を持つ多数の仮説群を用いて離散的に近似し、事前の状態と状態遷移モデルに基づく“予測”、観測画像を用いて人物らしさを評価する“観測”の処理を逐次的に処理することで実現される。動きの変動や観測のノイズ、特に観測値が非ガウス型になるような場合においても頑健な追跡を実現できることから、実環境下における追跡手法として広く利用されている^{1),5)-8),10)}。

しかしながら、上述した問題から実環境において追跡の失敗を防ぐことは難しい。たとえば、追跡している人物が障害物によって完全に遮蔽され、そして遮蔽された場所とは異なる位置から再度現れる場合、システムは追跡していた人物を見失ってしまうため、正しい追跡をただちに再開することは困難である。このような問題に対処するために、追跡が不安定になる、もしくは失敗した際に、修正、再初期化を速やかに行うことで正しい追跡を再開する手法がいくつか提案されている^{5),6),8)}。これらの手法は、観測画像から得られる指標のうち、追跡に利用する指標とは異なるものを追跡の初期化用として利用することで、追跡失敗の問題に対処している。

一方、鈴木らは、画像から得られる指標に加え、環境モデルと呼ばれる対象空間内の人物の存在確率分布を導入することで追跡の安定化を図る手法を提案している¹⁰⁾。彼らは、対象空間内の机や椅子などの物理的障害物からの距離や配置に基づき環境モデルを定義している。これにより、壁、床の下など人物が存在しえない領域に仮説が発生することを抑制す

^{†1} 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo

^{†2} 国立情報学研究所
National Institute of Informatics

*1 現在、埼玉大学
Presently with Saitama University

ることができるため、安定な追跡が実現される。しかしながら、この手法はあらかじめ環境モデルが獲得されていない場合には使用できないという問題がある。また、環境の物理的な制約だけでは対象空間内の人物の存在確率分布を必ずしも表現できていない。

たとえば、対象空間内に2つの通路があり、そのうち1つの通路のみが頻繁に使われるシーンを考える。この場合、物理的な制約のみに基づいて人物存在確率を表現すると、両者はほぼ等しい確率となってしまうが、実際には人物が頻繁に通行する通路の方が人物の存在確率は高いと考えられる。これは、人物の存在確率は物理的制約よりも対象空間中の人物の行動履歴に強く依存することを示唆している。このような人物の行動の履歴に着目し、それを利用することで、対象空間に特化した人物存在確率分布を表現することが可能となる。そしてこれを人物追跡の初期化のための情報源として利用することで、観測環境に適応した高速、高精度な追跡の再初期化を実現することができると考えられる。

このような知見に基づき、本稿では、人物の行動履歴に基づいて対象空間の人物存在確率分布を獲得し、追跡の枠組みに組み込むことで、人物追跡の安定化を図る手法を提案する。以後本稿では、このような人物存在確率分布を“環境属性”と呼ぶ。提案手法は具体的には次のように実現される。ICONDENSATION⁵⁾の考えに基づき、環境属性をimportance functionとしてパーティクルフィルタの追跡の枠組みに組み込む。また環境属性は混合正規分布でモデル化し、オンラインEMアルゴリズム⁹⁾を用いることで、追跡器により毎フレーム推定される人物頭部位置から逐次的に学習させる。さらに、人物の行動状態推定を導入し、行動状態に応じて環境属性の学習の重みを適応的に制御する。これより、対象シーンに特化した適応的な環境属性の獲得と追跡の安定化を同時に実現する。

本手法の応用目的は次のように考えられる。環境属性は時間と場所に応じた人物存在確率を表す。そのため、これは店内のレイアウト設計のための指標、もしくは駅構内における混雑解消のための情報源として利用できると考えられる。

不審行動検出などの上位のアプリケーションのための基礎技術としての利用も考えられる。人物存在確率が高い領域においては、環境属性により高速かつ安定な追跡初期化が働くため、安定した追跡軌跡が得られる。そのため直接上位アプリケーションへ情報を供給することが可能となる。人物存在確率が低い領域においては、間接的に上位アプリケーションに寄与できると考えられる。人物存在確率が低い領域は、人物が存在しにくい領域であることを示している。そこで、このような人物存在確率の偏りを他システムへ通知し、その領域を集中的に監視させることにより、より効果的に不審者検出を行うことができると考えられる。

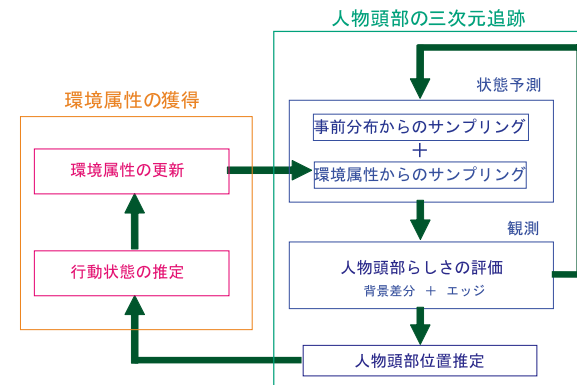


図1 システム概要
Fig. 1 System overview.

本稿の構成は以下のとおりである。まず、2章において環境属性の人物追跡への統合方法について述べる。続く3章では、環境属性の構築法について詳しく述べる。そして4章では本手法の有効性、適用範囲について実験を通して考察する。最後に5章で結論を述べる。

2. 環境属性の人物頭部追跡への統合

図1に本システムの全体図を示す。本手法は“人物頭部の三次元追跡”と“環境属性の獲得”の2つの処理単位が相互に作用することで実現される。これにより人物追跡の安定化と環境属性の獲得を同時に実現することが可能となる。以下、本章では、“人物頭部の三次元追跡”の処理単位について述べる。

2.1 パーティクルフィルタによる人物追跡

時刻 t における対象の状態量を X_t 、時刻 t までに得られた画像による観測列を $Z_t = \{z_1, \dots, z_t\}$ とする。時系列フィルタによる人物追跡は、追跡対象を表す密度分布 $p(X_t | Z_t)$ を逐次推定することであり、それは、追跡対象の事前の状態に基づく“予測”と、観測画像を用いて人物らしさを評価する“観測”のステップを逐次実行することで実現される。

パーティクルフィルタは、密度分布 $p(X_t | Z_t)$ を重み $\pi_t^{(n)}$ を持つ仮説 $s_t^{(n)}$ の集合 $\{(s_t^{(n)}, \pi_t^{(n)}) \mid n = 1, 2, \dots, N\}$ を用いて離散的に近似する。これにより、任意の形状の分布を表現することが可能となるため、ノイズや環境変動に対して頑健な追跡を実現できる。具体的には次の処理を繰り返すことで実現される。

(i) 仮説の選択

時刻 $t - 1$ における事後分布 $p(\mathbf{X}_{t-1} | \mathcal{Z}_{t-1})$ を表す仮説群 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ が持つ重み $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ の比に従い、仮説群 $\{s_{t-1}'^{(1)}, \dots, s_{t-1}'^{(N)}\}$ を選択する。

(ii) 状態遷移モデルに基づく仮説の伝播

選択された仮説群 $\{s_{t-1}'^{(1)}, \dots, s_{t-1}'^{(N)}\}$ を、あらかじめ定義した状態遷移モデル $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ に従いそれぞれ伝播させ、時刻 t における事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ に従う仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ を生成する。

(iii) 観測による重み $\pi_t^{(n)}$ の決定

仮説 $s_t^{(n)}$ の重み $\pi_t^{(n)}$ を、時刻 t の観測画像から得られる観測モデル $p(\mathbf{Z}_t | \mathbf{X}_t)$ を用いて評価することにより決定する。

本手法の人物追跡器は、視野を共有する複数台のカメラを用いて人物頭部を三次元的に追跡する。室内空間において、床面を XY 平面に一致させ、高さ方向に Z 軸をとる三次元世界座標系 $O - XYZ$ を定義する。追跡対象である人物頭部を楕円体でモデル化し、その中心座標を (x, y, z) とする。また時刻 t における n 番目の仮説は、 $s_t^{(n)} = [x_t^{(n)}, y_t^{(n)}, z_t^{(n)}]^T$ という三次元ベクトルの状態量を持つものとする。

2.2 環境属性の統合

環境属性を importance function としてパーティクルフィルタの追跡の枠組みに導入する。すなわち、仮説をサンプリングするとき、事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ と環境属性 $g_t(\mathbf{X}_t)$ をサンプリング源として同時に利用する。具体的には、ICONDENSATION⁵⁾ の考え方にに基づき以下の3つのサンプリング手法を併用する。またこれらのサンプリング手法はある一定の割合で選択する。

(a) 追跡：パーティクルフィルタの枠組みにおける仮説のサンプリング。事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ からサンプリングする。

(b) 追跡初期化：追跡初期化のための importance sampling。環境属性 $g_t(\mathbf{X}_t)$ からサンプリングする。

(c) 追跡修正：追跡修正のための importance sampling。環境属性 $g_t(\mathbf{X}_t)$ からサンプリングする。仮説の重みを評価するときに環境属性により修正する。

これにより環境に適応した追跡の修正、初期化を行うことが可能となるため、より安定な追跡を実現することができると考えられる。環境属性の詳細については3章で述べる。

2.3 人物頭部らしさの評価

各カメラで観測される画像の特徴量を用いて、仮説 $s_t^{(n)}$ の人物頭部らしさを評価する。具

体的には、仮説をカメラに投影したときの画像において、背景差分により抽出される前景領域の割合と、エッジ勾配の類似性を用いて、その仮説の頭部らしさを評価する。 i 番目のカメラの背景差分による評価値を $\pi_{i,t}^{fg,(n)}$ 、 i 番目のカメラのエッジ勾配の類似性に基づく評価値を $\pi_{i,t}^{edge,(n)}$ としたとき、各カメラで評価される人物頭部らしさ $\pi_{i,t}^{(n)}$ を、これらの積により与える。そして各々のカメラについて評価した後、それぞれの重みの積をとることで、その時刻における人物頭部らしさの評価値 $\pi_t^{(n)}$ とする。なお、サンプリング手法 (c) より発生した仮説の重みは、次のように定義する。

$$\pi_t^{(n)} = \frac{p(\mathbf{X}_t = s_t^{(n)} | \mathcal{Z}_{t-1})}{g_t(s_t^{(n)})} \prod_i \pi_{i,t}^{(n)}. \quad (1)$$

ここで $g_t(\mathbf{X}_t)$ は importance function であり、環境属性を意味する。すべての仮説に対して重みを計算した後、事後分布の期待値をとることにより、時刻 t における人物頭部の三次元位置 s_t^{head} を推定する。

3. 環境属性の更新

3.1 環境属性の確率表現

複雑な環境下での広い空間を対象とする場合、人物の存在しやすい領域は複数あると考えられる。このような環境における人物存在確率を表現するために、環境属性は多峰性の分布に対応できる必要がある。また、前述したように環境属性を importance function $g_t(\mathbf{X}_t)$ として追跡の枠組みに組み込むため、仮説のサンプリングが容易なモデルである必要がある。そこで本手法では、環境属性を混合正規分布を用いて表現する。すなわち、

$$g_t(\mathbf{X}_t) = \sum_{i=1}^K \omega_{i,t} \mathcal{N}(\mathbf{X}_t | \boldsymbol{\mu}_{i,t}, \boldsymbol{\Sigma}_{i,t}). \quad (2)$$

ここで $\mathcal{N}(\cdot)$ は正規分布を表し、 $\boldsymbol{\mu}_{i,t}$ 、 $\boldsymbol{\Sigma}_{i,t}$ はそれぞれ時刻 t における i 番目の正規分布の平均、共分散を意味する。また $\omega_{i,t}$ は時刻 t における i 番目の正規分布の重み、そして K は正規分布の数を表す。このようなパラメトリックモデルで表現した環境属性に行動履歴を反映させるということは、行動履歴を用いてモデルが持つパラメータを更新することと同義である。

3.2 追跡結果の動的反映による環境属性の更新

追跡器により毎フレーム得られる人物頭部の三次元位置 s_t^{head} を環境属性に逐次反映させ

ることにより、行動履歴に基づく環境属性を獲得する。具体的には、時刻 $t-1$ において各正規分布が持つパラメータ $\omega_{i,t-1}, \mu_{i,t-1}, \Sigma_{i,t-1}$ と人物頭部の三次元位置 s_t^{head} を用いて、オンライン EM アルゴリズムを環境属性に逐次適用することで実現する。

EM アルゴリズムとは、最尤推定手法の 1 つであり、観測できない、もしくは未観測の事象が存在する場合においても用いることができる汎用的な最尤推定手法である²⁾。推定すべきパラメータを θ 、観測されるデータ集合を $\chi = \{x_1, x_2, \dots, x_m\}$ とし、観測できない値をとる確率変数を Y とする。このとき、EM アルゴリズムでは、以下の 2 つのステップを解が収束するまで繰り返す。

E ステップ: 次の対数尤度の条件付き期待値を計算する。

$$Q(\theta|\theta^{(t)}) = E [\ln p(\mathcal{X}, \mathbf{Y} | \theta) | \mathcal{X}, \theta^{(t)}]. \quad (3)$$

M ステップ: $Q(\theta | \theta^{(t)})$ を θ について最大化し、その θ を $\theta^{(t+1)}$ とする、すなわち、

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta | \theta^{(t)}). \quad (4)$$

このように、EM アルゴリズムは全観測データを用いたバッチ処理によって実現される。このため、多数の観測データを用いる場合、多くの計算量を必要とする。

一方で、EM アルゴリズムにおける計算量の問題を克服し、オンライン処理に拡張したオンライン EM アルゴリズムが提案されている⁹⁾。そこでは、時刻 t において得られた観測データ $x(t)$ と時刻 $t-1$ で推定されたパラメータを用い、時刻 t におけるパラメータを推定する。各混合正規分布のパラメータ $\omega_{i,t}, \mu_{i,t}, \Sigma_{i,t}$ は次のようにそれぞれ計算される。

$$\mu_i^{(t)} = \frac{\langle\langle \mathbf{x} \rangle\rangle_i(t)}{\langle\langle 1 \rangle\rangle_i(t)}. \quad (5)$$

$$\Sigma_i^{(t)} = \frac{\langle\langle \mathbf{x} \mathbf{x}^T \rangle\rangle_i(t)}{\langle\langle 1 \rangle\rangle_i(t)} - \mu_i^{(t)} \left(\mu_i^{(t)} \right)^T. \quad (6)$$

$$\omega_i^{(t)} = \langle\langle 1 \rangle\rangle_i(t). \quad (7)$$

ここで、 $\langle\langle \cdot \rangle\rangle_i(t)$ は、学習の重みを α として、

$$\langle\langle f(\mathbf{x}) \rangle\rangle_i(t) = (1 - \alpha)\langle\langle f(\mathbf{x}) \rangle\rangle_i(t-1) + \alpha f(\mathbf{x}) p(i | \mathbf{x}(t), \theta^{(t-1)}), \quad (8)$$

として定義される。また $p(i | \mathbf{x}(t), \theta^{(t-1)})$ は i 番目の正規分布が選択される事後分布を意味する。このように、過去の推定結果を学習率 α で伝播させることにより、高速なパラメータ推定を可能にしている。そのため、オンライン EM アルゴリズムを利用することは、人

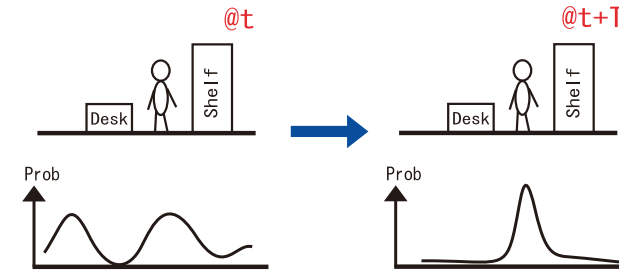


図 2 滞留行動による環境属性の局所化
Fig. 2 Localization of the EEM due to standing-still actions.

物頭部追跡と環境属性の獲得を同時に実現する本手法に適している。

3.3 人物の行動状態推定にともなう行動履歴の適応的反映

人物の行動により環境属性の本来の役割である追跡の安定化、初期化の機能が失われる場合がある。この問題を引き起こす人物の行動例の 1 つとして、滞留行動が考えられる。人物が滞留している間、人物頭部の位置はほとんど変化しない。長時間このような状態が続いたとき、行動履歴の環境属性への反映により、滞留している周辺位置にのみ環境属性の確率密度値が集中する状態が起こる(図 2)。これにより、他の領域は人物が存在しえない領域であると判断され、人物追跡の不安定化、新規追跡対象の検出失敗などといった様々な問題を引き起こす可能性がある。このような問題に対処するために、追跡対象である人物の行動状態を推定し、状態に応じて環境属性の更新の重み α を適応的に変化させる必要がある。そこで本手法では、図 3 に示す行動状態遷移モデルを導入することで、追跡対象の行動状態にともなう環境属性の適応的更新を実現する。

3.3.1 視野外/視野内の判定

追跡対象である人物が視野内に存在しているかどうかの判定条件を定義する。本手法では、この状態判定に追跡中における尤度の変化を利用する。もし人物が視野内に存在するならば、人物が存在する領域周辺で尤度が相対的に高くなることが考えられる。逆に人物が視野内に存在しないとき、いずれの領域においても人物らしさが高くなることはないため、尤度は一様に小さい値をとる。このような人物の有無にともなう尤度分布の変化を次のように利用する。

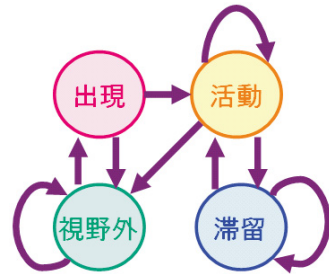


図 3 行動状態遷移モデル

Fig. 3 Human action state transition model.

視野外/視野内判定条件

- $\sum_{i=1}^N \pi_t^{(i)} < T_{exist}$: 視野外と判定 (1)
- $\sum_{i=1}^N \pi_t^{(i)} \geq T_{exist}$: 視野内と判定 (2)

ここで $\pi_t^{(i)}$ はパーティクルフィルタにおける i 番目の仮説が持つ尤度を意味し, N は全仮説数を意味する. また T_{exist} は判定のための閾値である.

3.3.2 行動状態遷移モデル

図 3 に示すように, 行動状態遷移モデルは視野外, 出現, 活動, 滞留の 4 つの状態を持つ. また人物はこのいずれかの状態に属すると仮定する. 以下それぞれの状態, 遷移条件について詳しく述べる.

視野外状態

人物がカメラにより観測されていない状態を視野外状態とする. このとき人物は視野内に存在しないため, 環境属性の更新の重みは $\alpha = 0$ とする. 視野外状態からの遷移条件は, 前節の視野外/視野内判定の方法を用いて次のように定義する.

- 条件 (1) を満たす: 遷移しない (視野外状態)
- 条件 (2) を満たす: 出現状態へ遷移

出現状態

視野外から人物が新たに現れた状態を出現状態とする. この状態は追跡初期化に重要であると考えられるため, 環境属性の更新の重みを $\alpha = k\alpha_{act}$ とする. ここで $k > 1$ は任意の定数, また α_{act} は活動状態における更新の重みを表す. 出現状態からの遷移条件を次のように定義する.

- 条件 (2) を満たす: 活動状態へ遷移

- 条件 (1) を満たす: 視野外状態へ遷移

滞留状態

ある特定の場所から人物が長時間ほぼ動かない状態を滞留状態とする. 環境属性の集中にともなう初期化の機能喪失を防ぐために, 滞留時間 t_{stay} の増加にともない滞留状態における環境属性の更新の重みを減少させる. 具体的には $\alpha = \alpha_{act}/t_{stay}$ とする. 滞留状態からの状態遷移について以下の 2 つの条件を用いる.

- 人物の移動速度 v_t が閾値 T_{vel} [cm/frame] 以下であることを満たす. (3)
- 条件 (3) を閾値 T_s [frame] 以上満たす. (4)

これらの条件を利用し, 滞留状態からの遷移条件を次のように定義する.

- 条件 (2), (3) を満たす: 遷移しない (滞留状態)
- 条件 (2) を満たす: 活動状態へ遷移

活動状態

視野内で人物が何らかの行動をしている状態を活動状態とする. 多くの場合人物はこの状態に属する. 環境属性の更新の重みは $\alpha = \alpha_{act}$ とし, 活動状態からの遷移条件は次のように定義する.

- 条件 (2) を満たす: 遷移しない (活動状態)
- 条件 (2), (4) を満たす: 滞留状態へ遷移
- 条件 (1) を満たす: 視野外状態へ遷移

4. 実 験

本手法の有効性について確認するために, 図 4 に示すような対象空間において, 環境属性の獲得実験と人物頭部追跡実験を行った. 室内天井に校正済みの IEEE1394 カメラ (Point Grey Research 社製 Flea) を 2 台設置し, 解像度 640×480 ピクセル, 毎秒 30 フレームの速度で得られるカメラ画像を用いた. カメラ 1 台ごとに 1 台の汎用 PC (CPU: Intel Pentium4 2.8 GHz, RAM: 2 GByte, OS: WindowsXP) を割り当て, それぞれをクライアント PC とした. またクライアント PC の情報を統合するサーバの役割を果たす汎用 PC (CPU: Intel Pentium4 3.2 GHz, RAM: 1 GByte, OS: WindowsXP) を用意し, これらを通信速度 1 Gbps の Gigabit Ethernet で接続してサーバ・クライアントモデルを構築した. 実験で用いたパラメータは次のように設定した. 環境属性の正規分布の個数は $K = 7$ とし, 行動状態推定に用いたパラメータはそれぞれ $\alpha_{act} = 0.0005$, $k = 2$, $T_{exist} = 2.5$, $T_{vel} = 1$ [cm/frame], $T_s = 200$ [frame] とした. また, Mixture Particle Filter¹¹⁾ の考え

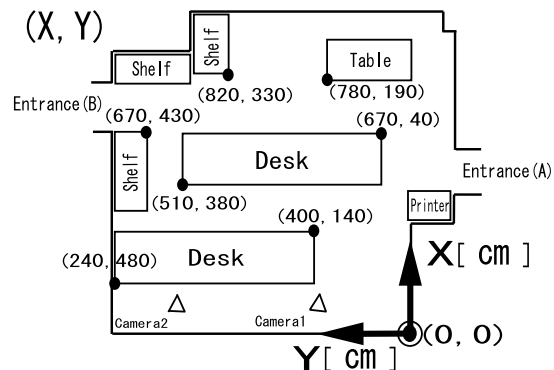


図 4 対象空間のレイアウト：図中の数値はそれぞれの位置における座標値を意味する

Fig. 4 Layout of the observation space: Each pair of the numbers (x, y) in the figure represents the coordinate value on each point.

方に基づき、複数人の追跡の枠組みに拡張して実験を行った。人物 1 人あたりの追跡に利用した仮説数は $N = 150$ とした。

4.1 環境属性の獲得

対象空間をおよそ 30 分間観測することにより、環境属性の獲得実験を行った。本実験では、行動状態推定の効果を確認するために、行動状態推定をともなう場合とそうでない場合のアルゴリズムをあらかじめ撮影した映像に対してそれぞれ適用し、得られた分布を比較した。ここでは、片方の通路の頻繁な通行、特定個所における雑談による滞留が主にみられた部分の映像を利用した。図 5 にその様子を示す。

それぞれの推定結果を図 6、図 7 に示す。動線の反映のみにより得られた環境属性 (図 6) は、ある領域 ($650 < X < 750, 0 < Y < 100$) における確率密度値がほかの部分に比べ極端に高くなっていることが分かる。これは人物の滞留行動による環境属性の確率密度値の集中によるものと考えられ、実験に利用した映像においても 2 人の人物が長時間雑談している様子が観測された (図 5-(ii))。このことから、実際に観測された滞留行動が正しく反映されていることが確認できる。しかしながら、その一方で、滞留行動以外に頻繁に観測された片方の通路の歩行による反映の効果がみられない。これは、式 (8) における学習率を一定の値として行動履歴を反映し続けたことに起因すると考えられる。行動状態推定を利用することにより得られた環境属性 (図 7) では、前者の環境属性においても同様の領域でみられた分布のピーク、そして奥側の通路領域 ($650 < X < 750, 0 < Y < 450$) に沿ったなだ



(i) 特定の通路の歩行

(ii) 雑談による滞留

(i) Walking at a particular walkway

(ii) Standing still for chatting

図 5 頻繁に観測された行動

Fig. 5 Frequently observed human actions.

らかな分布の峰がみられる。これは、片側の通路の頻繁な通行 (図 5-(i)) によるものと考えられる。このことから行動履歴が正しく反映されていると見なすことができる。これは、人物の行動状態にともない行動履歴が適応的に環境属性へ反映されたことによる効果であると考えられる。

以上のことから、人物の行動履歴が正しく環境属性に反映されていることが確認することができた。また行動状態推定は行動履歴を正しく反映させるために必要であることを確認することができた。

4.2 人物追跡における環境属性の有用性

4.1 節で構築された行動状態推定をともなう環境属性 (図 7) を用いて、追跡初期化に関する環境属性の有効性を確認した。

4.2.1 環境属性による追跡初期化の高速性

人物が視野内に現れたとき、もしくは遮蔽などなんらかの要因によって追跡に失敗した状態から正しい追跡が再開されるまでの早さを評価した。追跡対象が実際に観測されるフレームを f_{gt} 、正しく追跡が初期化されたときのフレームを f_{init} とする。このとき両者のフレーム差、 $f_{diff} \equiv |f_{gt} - f_{init}|$ を追跡初期化が発生する様々な場合について計算し、統計的に評価した。これは、 f_{diff} が 0 に近いほど高速に追跡初期化が行われていることを意味する。図 8 に例を示す。

本手法の有効性を比較検証するために、行動履歴を反映することにより得られた環境属性

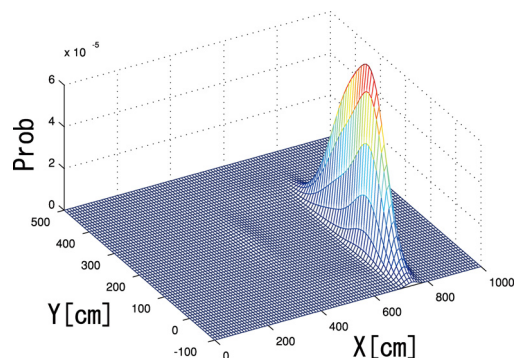


図 6 環境属性の構築結果（行動状態推定なし）

Fig. 6 Result of the constructed EEM without human action state transition model.

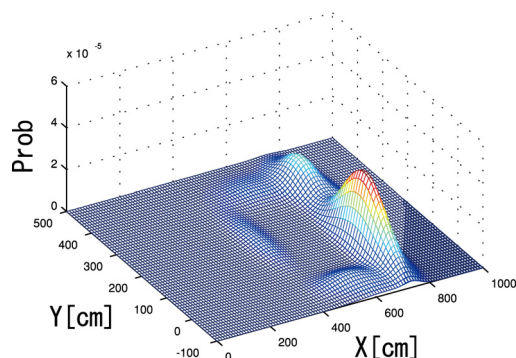
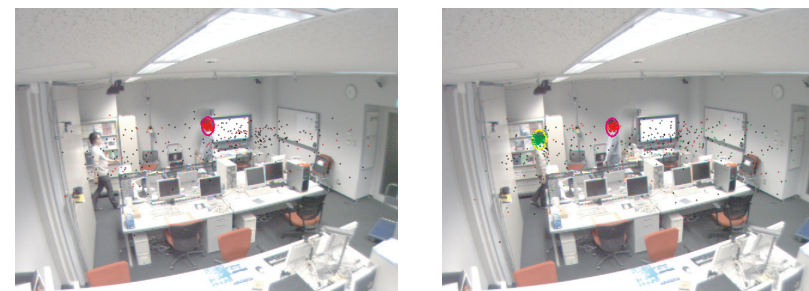


図 7 環境属性の構築結果（行動状態推定あり）

Fig. 7 Result of the constructed EEM with human action state transition model.

を用いた場合と一様分布を環境属性として利用した場合の両者を用いて、単一シーンにおいて比較実験を行った。本実験では、人物は得られた環境属性に従うような行動、すなわち人物存在確率が高い領域において頻繁に行動する映像を用いた。また本映像では、人物同士のすれ違いによる部分遮蔽、人物が机下に入り込むことによる完全遮蔽などが頻繁に観測された。試行回数は 5 回であり、追跡の初期化、再初期化は全試行合わせて 40 回行われた。

初期化の速さの比較結果を表 1 に示す。示されている値はそれぞれ f_{diff} の平均、標準偏差、すなわち追跡が初期化されるまでに必要な平均フレーム数とそのばらつき具合を意味す



(a) # 1038 : 新規追跡対象出現

(b) # 1054 : 追跡開始

(a) # 1038: A new person appears

(b) # 1054: Tracking is started

図 8 検証実験評価例：# 1038 (左図 (a)) では人物が新たに現れたことが観測できるが、追跡は開始されていない。そして # 1054 (右図 (b)) で追跡が開始されている。この例では、# 1038 を追跡が開始されるべき真値 f_{gt} 、# 1054 を実際に追跡開始されたフレーム f_{init} とする。そしてこれらの差分 $f_{diff} = |f_{gt} - f_{init}|$ を計算する

Fig. 8 Example of the verification experiment: In #1038 (a), a new person appeared but was not tracked yet. In #1054 (b), our tracker started tracking the person. We set the #1038 f_{gt} and the #1054 f_{init} , then calculate $f_{diff} = |f_{gt} - f_{init}|$.

表 1 追跡初期化の速さの比較結果：人物が実際に観測されたフレームと正しい追跡初期化が働いたフレームとの差、 f_{diff} の平均と標準偏差を意味する

Table 1 Comparison results of the initialization speed: the mean and the standard deviation are of f_{diff} .

	平均 [frame]	標準偏差 [frame]
一様分布	48.31	37.92
環境属性	21.62	16.45

る。環境属性と一様分布それぞれの平均を比較すると、環境属性によるものは一様分布のそれに比しておよそ 1/2 程度のフレーム数で追跡初期化が行われている。この環境属性による平均遅れは時間にして 1 秒弱の遅れであり、高速な追跡初期化が実現されているといえる。また標準偏差の比較結果について着目すると、環境属性によるものは一様分布のそれに比しておよそ 1/2 程度に抑えられており、安定な初期化が実現できていることが分かる。

4.2.2 人物頭部追跡結果

環境属性と一様分布を利用した場合の追跡精度の比較

人物 2 人 (人物 A, 人物 B) の頭部の追跡実験を単一シーンにおいて行った。環境属性の

表 2 追跡誤差
Table 2 Comparison results of tracking errors.

		平均誤差 [cm]		標準偏差 [cm]	
		一様分布	環境属性	一様分布	環境属性
人物 A	XY 平面	15.75	14.54	10.06	9.10
	Z 軸	4.98	4.55	5.53	5.23
人物 B	XY 平面	15.18	13.52	8.23	7.32
	Z 軸	5.22	3.79	6.18	4.08

効果を比較検証するために、環境属性、一様分布をそれぞれ利用して実験を行った。人物 A は出入口 (A) から視野内に現れ、別の出入口 (B) に向かって歩いた。それと同時に、人物 B は出入口 (B) から視野内に現れ、人物 A が歩いた通路と同じ通路を歩き、部屋の奥の方へ歩いた。また追跡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することで求めた三次元座標を真の位置と見なした。試行回数は 5 回であり、それぞれのシーケンスについて追跡結果と真値の比較を行った。

全試行について統計的に評価した追跡誤差結果を表 2 に示す。それぞれ Z 軸方向および XY 平面上での平均誤差と標準偏差を表す。環境属性を利用した場合、一様分布を利用した場合の評価結果を比較すると、両者ほぼ同等の追跡精度が得られていることが分かる。XY 平面上の平均誤差が 15 cm 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 5 cm 程度に対応することを考慮すると、安定な追跡を実現できているといえる。あるシーケンスにおける人物追跡結果を図 9 に、追跡の様子を図 10、図 11 に示す。これらの様子からも両者ともに安定な追跡を実現できていることが見てとれる。

遮蔽への対処

図 12 は、人物が奥の通路を歩き (# 345)、途中設置されたボードにより完全に遮蔽された後 (# 360)、ボードの逆側から現れ (# 720)、奥の部屋へ向かうシーケンスに対する追跡結果の一部である。この結果にあるように、ボードの完全な遮蔽により追跡が失敗していることが確認できる (# 360 から # 719)。しかしながら、その人物がボードの逆側から現れた際 (# 720)、追跡器は速やかに追跡を再開していることが分かる (# 723)。このように、環境属性を利用することで完全な遮蔽により追跡対象を見失った場合においても正しい追跡を再開できることが分かる。

以上の検証実験の結果から、環境属性は追跡の安定化、特に追跡対象の初期化、再初期化に関して効果的であるといえる。

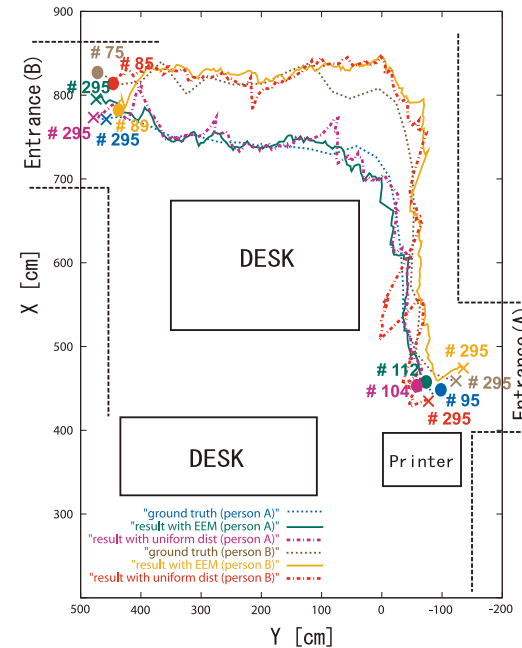


図 9 人物頭部追跡結果
Fig. 9 Tracking results of two human heads.

4.3 環境属性の更新に要する計算コスト

環境属性の更新は、環境属性を利用しない場合（一様分布を利用する場合）に比べていくらかの計算コストを要する。一方、パーティクルフィルタを用いた追跡において、仮説数の増加にともない追跡精度が向上することが知られている。これらのことから、環境属性の更新に用いる計算コストをパーティクルフィルタの仮説数の増加に用いることで、追跡の安定化を図ることができるとも考えられる。このような計算コストの面から本手法の有用性を検証するために、環境属性の更新に必要な計算コストを実験的に求めた。

4.2.1 項の実験で利用した映像に対して本手法を適用し、環境属性の更新に要した処理時間を各フレームごとに計測した。そしてその処理時間の平均値を環境属性の更新に必要な処理時間として採用した。4 章冒頭で述べた実験環境で計測した結果、1 回の環境属性の更新に必要な処理時間はおよそ 0.0025 [ms] であった。



図 10 2人の人物頭部追跡の様子（一様分布の利用）

Fig. 10 Tracking result of two human heads using the uniform distribution.



図 11 2人の人物頭部追跡の様子（環境属性の利用）

Fig. 11 Tracking result of two human heads using the EEM.

次に、環境属性の更新を行わない場合（一様分布の利用）において、1 仮説を処理するために必要な処理時間を同様の方法で計測した。計測した結果、1 仮説を処理するために必要な処理時間はおよそ 0.0211 [ms] であった。これらの計測結果を用いて、追加可能な仮説の数 N' を算出すると、仮説の数はおよそ 0.0946 [個] となる。これはすなわち、環境属性の更新に必要な計算コストを利用しても仮説の追加を行うことはできないことを意味する。このことから、環境属性の更新はパーティクルフィルタの仮説を増加させることに比べて低い計算コストで行うことができることを確認できた。また 4.2.1 項の実験結果とあわせて考えると、環境属性を利用することで仮説の数を増加させることなく追跡の安定化を図ることができるといえる。

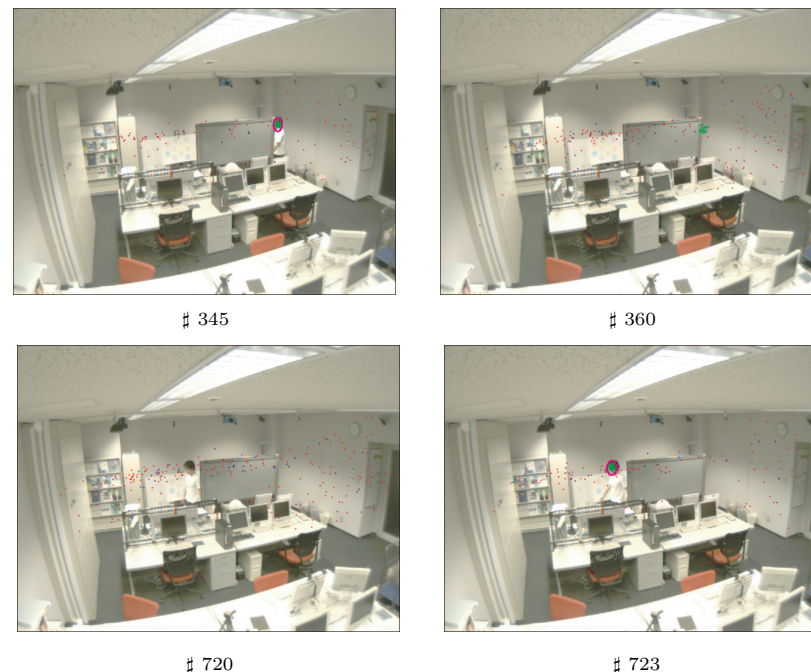


図 12 追跡結果：完全な遮蔽からの追跡再開

Fig. 12 Tracking result: Recovering tracking from complete occlusions.

4.4 環境属性の適用範囲

観測履歴と異なり、環境属性の人物存在確率が低い領域において人物が頻繁に行動する場合、環境属性を利用することにより逆に追跡が不安定化することが考えられる。環境属性の適用範囲を知るために、本実験では環境属性が不得手とする場合について検証を行った。具体的には、4.1 節で得られた環境属性において人物存在確率が低い領域 ($400 < X < 500$, $0 < Y < 400$) を頻繁に人物が行動する状況と考えた。また 4.2 節の実験結果から分かるように、追跡安定化に関する環境属性の寄与は主に追跡初期化にあることから、4.2.1 項と同様の追跡初期化の速さに関する評価を行った。実験に用いた映像は、上述したように人物存在確率が低い領域において人物が頻繁に行動するものとした。その映像では人物同士のすれ違いによる部分遮蔽、柱による遮蔽、人物が机下に入り込むことによる遮蔽などが頻繁に観測された。試行回数は 5 回であり、追跡の初期化、再初期化は全試行合わせて 35 回行わ

表 3 人物存在確率が低い領域における追跡初期化の速さの比較結果

Table 3 Comparison results of the initialization speed in the low probability area.

	平均 [frame]	標準偏差 [frame]
一様分布	36.22	38.78
環境属性	51.75	40.25

れた。

初期化の速さの比較結果を表 3 に示す。両者の結果を比較すると、一様分布を用いた方が高速かつ安定に追跡初期化が行われている。これは、追跡失敗が人物存在確率が低い領域のみで頻繁に起こったことにより、いずれの追跡初期化においても多くの時間を要したためであると考えられる。また 4.2.1 項の実験結果 (表 1) と比較すると、一様分布を用いた場合は平均、標準偏差ともに結果に大きな差はないといえる。それに対し環境属性を用いた結果では、人物が環境属性に従う行動をとった場合 (表 1) に比べ平均、標準偏差両面において明らかに不安定になっていることを確認できる。

これらの比較結果より、人物が観測履歴と大きく異なる動きをとる場合、環境属性は逆に追跡を不安定化させる可能性があるといえる。このような問題を防ぐために、より長時間の観測を対象とし、人物存在確率の極端な低下を防ぐ必要がある。また、環境属性だけでなく、他の異なる枠組みと併用する必要があると考えられる。

5. おわりに

人物は特定の通路を何度も通ったり、特定の位置で滞留したりするといったなんらかの行動の特性を持つことから、対象空間中の人物の存在確率はこのような行動特性に大きく依存すると考えられる。本稿ではこの点に着目し、人物の行動履歴を逐次利用することにより、対象シーンに特化した人物の存在確率を獲得できると考えた。そしてこれを環境属性として定義し、その構築法、追跡の枠組みへの統合法について提案した。また実データを用いた実験により、本提案手法の有効性を確認した。

本稿では、環境属性を混合正規分布で表現し、その正規分布の個数は経験的に定めている。そのため、個数の不足による追跡の修正、初期化の失敗の問題が発生することが考えられる。今後は、対象シーンに動的に正規分布の個数を対応させる方法について検討予定である。また、混合正規分布ではなく、パルツェン窓³⁾のようなノンパラメトリックな密度推定手法を用いて環境属性を表現する方法についても検討する予定である。

参考文献

- 1) Yang, C.R.D. and Davis, L.: Fast Multiple Object Tracking via a Hierarchical Particle Filter, *IEEE Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.212–219 (2005).
- 2) Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum-likelihood from incomplete data via the EM algorithm, *Journal of Royal Statistical Society B*, Vol.39, pp.1–22 (1977).
- 3) Duda, R.O., Hart, P.E. and Stork, D.G.: *Pattern Classification*, John Wiley & Sons, Inc. (2001).
- 4) 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).
- 5) Isard, M. and Blake, A.: ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework, *European Conference on Computer Vision*, Vol.1, pp.893–908 (1998).
- 6) Jin, Y. and Mokhtarian, F.: Data Fusion for Robust Head Tracking by Particles, *IEEE workshop on Visual Surveillance and Performance Evaluation and Tracking*, pp.33–40 (2005).
- 7) Kobayashi, Y., Sugimura, D., Sato, Y., Hirasawa, K., Suzuki, N., Kage, H. and Sugimoto, A.: 3D Head Tracking using the Particle Filter with Cascaded Classifiers, *the British Machine Vision Conference*, pp.37–46 (2006).
- 8) Okuma, K., Taleghani, A., Freitas, N., Little, J. and Lowe, D.: A Boosted Particle Filter: Multitarget Detection and Tracking, *European Conference on Computer Vision*, Vol.3021, pp.28–39 (2004).
- 9) Sato, M. and Ishii, S.: On-line EM Algorithm for the Normalized Gaussian Network, *Neural Computation*, Vol.12, no.2, pp.407–432 (2000).
- 10) 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 D-II*, Vol.J88-D-II, No.8, pp.1592–1600 (2005).
- 11) Vermaak, J., Doucet, A. and Perez, P.: Maintaining Multi-Modality through Mixture Tracking, *IEEE International Conference on Computer Vision*, Vol.2, pp.1110–1116 (2003).

(平成 19 年 9 月 24 日受付)

(平成 20 年 3 月 10 日採録)

(担当編集委員 日浦 慎作)



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、同大学院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。



小林 貴訓 (正会員)

2000～2004年三菱電機(株)設計システム技術センターにて、ソフトウェア生産技術の開発に従事。2007年東京大学大学院情報理工学系研究科博士課程修了。博士(情報理工学)。現在埼玉大学理工学研究科助教。コンピュータビジョン、ヒューマン・コンピュータ・インタラクションに関する研究に従事。



佐藤 洋一 (正会員)

1990年東京大学工学部機械工学科卒業。1997年カーネギーメロン大学大学院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。2008年電子情報通信学会論文賞、2006年電子情報通信学会論文賞、1999年情報処理学会山下記念研究賞、1999年日本バーチャルリアリティ学会論文賞等を受賞。電子情報通信学会、日本バーチャルリアリティ学会、ACM、IEEE各会員。



杉本 晃宏 (正会員)

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。2006～2007年Paris-Est大学客員教授。博士(工学)。視覚情報処理や離散システム・アルゴリズム等に広く興味を持ち、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。IEEE、ACM、日本応用数理学会、電子情報通信学会等の会員。