

行動履歴に基づいた環境属性の自動構築を伴う三次元人物追跡

杉村 大輔[†] 小林 貴訓[†] 佐藤 洋一[†] 杉本 晃宏^{††}

[†] 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: †{sugimura,yosinori,ysato}@iis.u-tokyo.ac.jp, ††sugimoto@nii.ac.jp

あらまし 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。

キーワード 人物追跡, 追跡初期化, 行動履歴, 環境属性, パーティクルフィルタ

Tracking People using Adaptive Environmental Attribute based on Long-Term Observation of Human Actions

Daisuke SUGIMURA[†], Yoshinori KOBAYASHI[†], Yoichi SATO[†], and Akihiro SUGIMOTO^{††}

[†] Institute of Industrial Science, The University of Tokyo Komaba 4-6-1, Meguro-ku, Tokyo, 153-8505 Japan

^{††} National Institute of Informatics Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: †{sugimura,yosinori,ysato}@iis.u-tokyo.ac.jp, ††sugimoto@nii.ac.jp

Abstract 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 obtain a probability of a person existing at a certain location. In this paper, we incorporate this probability as an importance function into the framework of particle filters. We call this probability the environmental attribute. The environmental attribute is iteratively updated at every frame by using the tracking results from our tracker. Our experimental results demonstrate the effectiveness of our method.

Key words people tracking, tracking initialization, human activity history, environmental attribute, particle filters

1. はじめに

コンピュータの高速化やカメラの小型化、低価格化に伴い、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 [5-8, 10, 12]。

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。例えば、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、正しい追跡を直ちに再開することは困難である。このような問題へ対処するために、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]。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、この手法は予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とも言えない。

例えば、対象空間内に二つの通路があり、そのうち一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。ICONDENSATION [5]の考え方に基づき、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する。そしてオンライン 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。更に、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。

2. 環境属性の人物頭部追跡への統合

図 2.1 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”の二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、本節では、“人物頭部の三次元追跡”の処理単位について述べる。

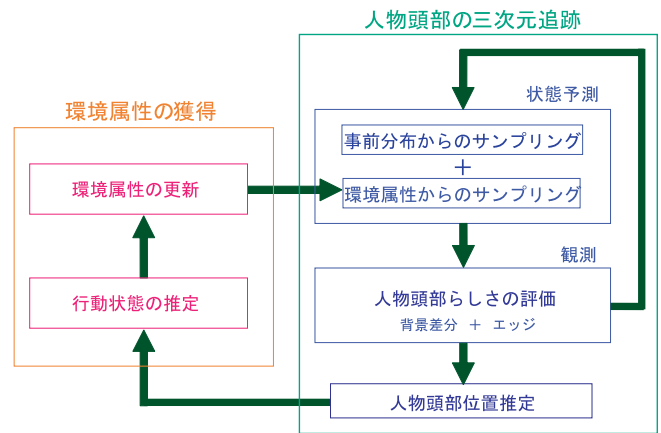


図 2.1: システム概要

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 \mathbf{X}_t 、時刻 t までに得られた画像による観測列を $\mathcal{Z}_t = \{z_1, \dots, z_t\}$ 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 $p(\mathbf{X}_t | \mathcal{Z}_t)$ 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 $p(\mathbf{X}_t | \mathcal{Z}_t)$ を重み $\pi_t^{(n)}$ を持つ仮説 $s_t^{(n)}$ の集合 $\{(s_t^{(n)}, \pi_t^{(n)}) | n = 1, 2, \dots, N\}$ 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、予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 $p(\mathcal{Z}_t | \mathbf{X}_t)$ 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、ICONDENSATION [5]の考え方に基づき以下の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節で述べる。

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。

$$\pi_t^{(n)} = \frac{p(\mathbf{X}_t = s_t^{(n)} | \mathcal{Z}_{t-1})}{g_t(s_t^{(n)})} \prod_i \pi_{i,t}^{(n)}. \quad (2.1)$$

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、

$$g_t(\mathbf{X}_t) = \sum_{i=1}^K \omega_{i,t} \mathcal{N}(\mathbf{X}_t | \boldsymbol{\mu}_{i,t}, \Sigma_{i,t}). \quad (3.1)$$

ここで $\mathcal{N}(\cdot)$ は正規分布を表し、 $\boldsymbol{\mu}_{i,t}$ 、 $\Sigma_{i,t}$ 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するという事は、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、 $\omega_{i,t-1}$ 、 $\boldsymbol{\mu}_{i,t-1}$ 、 $\Sigma_{i,t-1}$ 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、最尤推定手法の一つであり、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 [1]。推定すべきパラメータを θ 、観測されるデータ集合を $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ とし、観測できない値をとる確率変数を \mathbf{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期待値を計算する。

$$Q(\theta | \theta^{(t)}) = E \left[\ln p(\mathcal{X}, \mathbf{Y} | \theta) | \mathcal{X}, \theta^{(t)} \right]. \quad (3.2)$$

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、

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta | \theta^{(t)}). \quad (3.3)$$

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 $\mathbf{x}(t)$ 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 $\omega_{i,t}$ 、 $\boldsymbol{\mu}_{i,t}$ 、 $\Sigma_{i,t}$ は次のようにそれぞれ計算される。

$$\boldsymbol{\mu}_i^{(t)} = \frac{\langle \langle \mathbf{x} \rangle \rangle_{i(t)}}{\langle \langle 1 \rangle \rangle_{i(t)}}. \quad (3.4)$$

$$\Sigma_i^{(t)} = \frac{\langle \langle \mathbf{x} \mathbf{x}^T \rangle \rangle_{i(t)}}{\langle \langle 1 \rangle \rangle_{i(t)}} - \boldsymbol{\mu}_i^{(t)} \left(\boldsymbol{\mu}_i^{(t)} \right)^T. \quad (3.5)$$

$$\omega_i^{(t)} = \langle \langle 1 \rangle \rangle_{i(t)}. \quad (3.6)$$

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

$$\begin{aligned} \langle \langle f(\mathbf{x}) \rangle \rangle_{i(t)} &= (1 - \alpha) \langle \langle f(\mathbf{x}) \rangle \rangle_{i(t-1)} \\ &\quad + \alpha f(\mathbf{x}) p(i | \mathbf{x}(t), \theta^{(t-1)}) \end{aligned} \quad (3.7)$$

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伴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一つとして、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

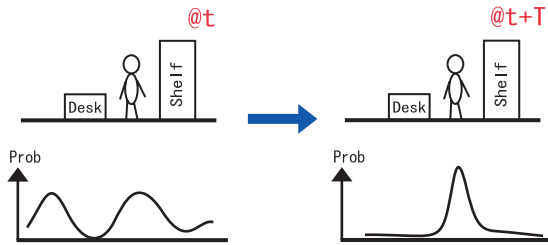


図 3.1: 滞留行動による環境属性の集中

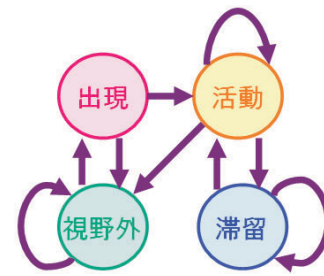


図 3.2: 行動状態遷移モデル

する状態が起こる (図 3.1). 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を適応的に変化させる必要がある. 本手法では, 簡単のために, 図 3.2 のような, 追跡において重要であると考えられる基本的な 4 つの行動状態を定義する. また, 人物はそれらの状態のうちいずれかに必ず属するものとする.

3.3.1 視野外/視野内の判定

図 3.2 の人物の行動状態モデルは, 視野外の状態とそれ以外の視野内の状態 (出現, 活動, 滞留) に分けることができる. 本手法では, この状態判定にパーティクルフィルタによる追跡の状態を示す次の指標 [3] を用いる.

$$N_{eff}(t) \equiv \frac{1}{\sum_{n=1}^N \{\pi_t^{(n)}\}^2}. \quad (3.8)$$

ここで $\pi_t^{(n)}$ は各仮説が持つ重みを表す. N_{eff} は追跡に有効な仮説の数を意味する. $N_{eff} \ll N$ であるとき数個の仮説のみが大きな重みを持つ縮退の状態であることを意味し, 逆に $N_{eff} = N$ であるとき全仮説の重みが均等であることを意味する. 具体的には, この指標を次のように利用する. 有効仮説数 N_{eff} がある一定数 T_v 以上で, かつ全仮説数 N よりも小さいとき, 視野内に人物頭部が存在することによる分布の偏りであるとみなすことができる. また $N_{eff} \sim N$ であるとき, 重みが均等であることから人物頭部は視野内には存在していないと判断できる. この判定方法を次のようにまとめる.

視野外 / 視野内 判定条件

- $N_{eff}(t) \sim N$: 視野外と判定 (1)
- $T_v < N_{eff}(t) < N$: 視野内と判定 (2)

3.3.2 行動状態とその遷移条件

図 3.2 の各行動状態における環境属性の更新の重みの設定, 状態遷移の条件についてそれぞれ述べる.

視野外状態

このとき人物は観測されていないため, 環境属性の更新の重みは $\alpha = 0$ とする. 視野外からの遷移条件は, 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)
- 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とき, 人物はこの状態に属する. この状態は追跡の初期化のため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は次のように定義する.

- 条件 (1) を満たす: 視野外へ遷移
- 条件 (2) を満たす: 活動状態へ遷移

活動状態

人物頭部が視野内にある (条件 (2)) とき, 多くの場合人物頭部はこの状態に属する. そのため更新の重みはこの状態にいるときを基本とし, その値を $\alpha = \alpha_{act}$ とする.

滞留状態との判別は, 以下の二つの条件を用いる.

- 空間的条件 (3): 異なる行動状態から活動状態へ遷移したときの人物頭部の初期三次元位置 s_0 と, 毎フレーム推定される頭部位置 s_t^{head} とのユークリッド距離が閾値 T_d [cm] 以下であることを満たす.
- 時間的条件 (4): 条件 (3) を閾値 T_s [frame] 以上満たす.

これより, 活動状態からの遷移条件を次のように定義する.

- 条件 (1) を満たす: 視野外へ遷移
- 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

滞留状態

ある一定の位置から長時間ほぼ動かない行動 (条件 (2),(4)) のときにこの状態に属する. 環境属性の集中という問題に対処するために, 滞留時間 t_{stay} の増加に伴い環境属性の更新の重みを減少させる. すなわち $\alpha = \alpha_{act}/t_{stay}$ とする. また滞留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: 活動状態へ遷移
- 条件 (2), (3) を満たす: 遷移しない (滞留状態)

4. 実験

本手法の有効性について確認するために, 図 4.1 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 x 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 台毎に 1 台の汎用 PC (CPU: Intel Pentium4 2.8GHz, RAM: 2GByte, 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 3.2GHz, RAM: 1GByte, OS: WindowsXP) を用意し, これらを通信速度 1Gbps の Gigabit Ethernet で接続してサーバ・ク

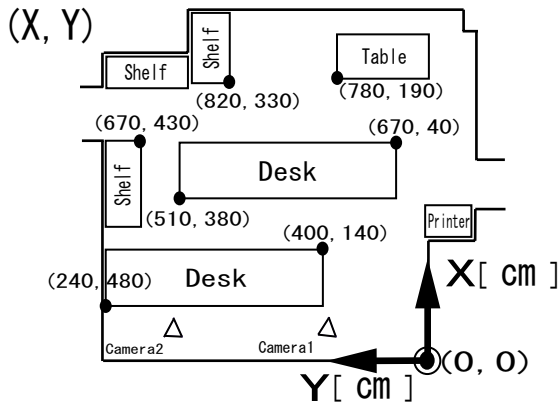


図 4.1: 対象空間のレイアウト: 図中の数値はそれぞれの位置における座標値を意味する。

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 $\alpha_{act} = 0.0005$, $k = 2$, $T_v = 30$ [個], $T_d = 50$ [cm], $T_s = 200$ [frame] とした。また, Mixture Particle Filter [11] の考え方にに基づき, 複数人の追跡の枠組みに拡張して実験を行った。人物 1 人当たりの追跡に利用した仮説数は $N = 200$ 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, 行動状態推定を伴う場合とそうでない場合のアルゴリズムを予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による滞留が主にみられた部分の映像を利用した。図 4.2 にその様子を示す。

それぞれの推定結果を図 4.3, 4.4 に示す。動線の反映のみ



(i) 特定の通路の歩行 (ii) 雑談による滞留

図 4.2: 頻繁に観測された行動

により得られた環境属性 (図 4.3) 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わかる。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二人の人物が長時間雑談している様子が観測された (図 4.2-(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 (3.7) 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 4.4) 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 4.2-(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みなすことができる。これは, 人物の行動状態に伴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がわかる。

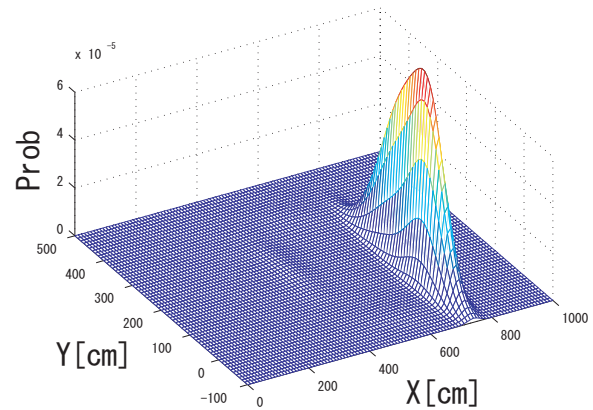


図 4.3: 環境属性 (行動状態推定無し)

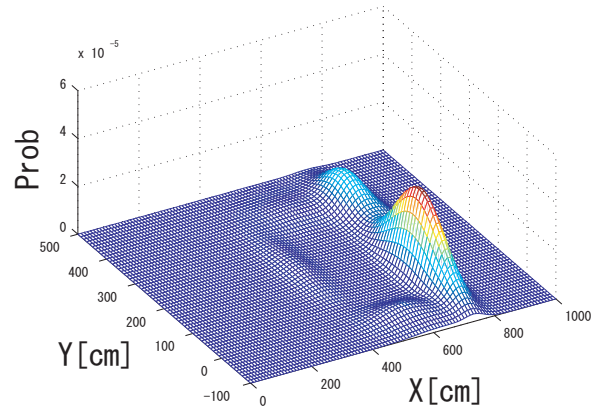


図 4.4: 環境属性 (行動状態推定有り)

4.2 人物追跡における環境属性の有用性

4.1 節にて構築された行動状態推定を伴う環境属性 (図 4.4) 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。具体的には, 高速な追跡初期化を実現できること, 遮蔽の問題に対処可能であることを確認した。

4.2.1 追跡初期化に関する環境属性の効果

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速さを評価した。追跡対象が実際に観測されるフレームを



1038: 新規追跡対象出現 # 1054: 追跡開始

図 4.5: 検証実験評価例: # 1038 (左図) 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) で追跡を開始している。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。 f_{diff} が 0 に近いほど高速な追跡初期化が行われることを意味する。

表 4.1: 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。

	平均 [frame]	標準偏差 [frame]
行動履歴に基づく環境属性	30.50	39.10
一様分布	87.90	72.70

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 4.5 に例を示す。また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において比較実験を行った。

初期化の速さの比較結果を表 4.1 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る。

行動履歴に基づく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 1/3 程度のフレーム数で追跡初期化を実現でき、これは時間にして約 1 秒程度の遅れで追跡を初期化できていることを意味する。このことから、環境属性を利用することで高速な追跡初期化を実現できていることがわかる。

次に標準偏差に着目すると、環境属性の利用による結果ではおよそ 40 フレーム程度のばらつきが見られる。これは、人物が存在しやすい領域 ($650 < X < 750$, $0 < Y < 450$) に比べて、人物がほとんど通らない手前側の通路 ($400 < X < 500$, $0 < Y < 450$) などといった確率密度値が小さい領域では追跡初期化にある程度の時間を要することに起因すると考えられる。しかしながらそれでも一様分布を利用した場合の半分程度の値であり、一様分布を利用するよりも安定に追跡初期化が実現されている。

追跡の比較結果を図 4.6, 4.7 に示す。環境属性を利用した結果 (図 4.7) は一様分布を利用した結果 (図 4.6) に比べて高速

な追跡初期化、安定な追跡が実現できていることが確認できる。以上より、環境属性を追跡初期化の情報源として利用することは有用であると言える。

4.2.2 遮蔽への対処

図 4.8, 図 4.9 に、障害物などによる静的な遮蔽、人物相互による動的な遮蔽について対処している様子をそれぞれ示す。図 4.8 は、人物が奥の通路を歩き (a)、途中ボードによる遮蔽により完全に隠れ (b)、そしてそのボードの逆側から現れて (c)、そのまま奥の部屋へ歩いて行く (d) というシーケンスに対する追跡結果の一部である。この結果からわかるように、ボードにより人物が遮蔽された ((b)) 後に異なる位置から出現しても ((c)), 直ちに正しい追跡を再開できていることがわかる ((c), (d))。

図 4.9 は、人物同士が通路ですれ違った際に、人物がもう 1 人の人物によって部分的に遮蔽されたときの追跡結果である。人物がすれ違うとき (e)、向かって奥の人物が手前の人物によって部分的に遮蔽されたことにより、その人物の追跡に失敗している (f)。しかしながら人物がお互いに離れた後 (g)、それぞれの人物頭部の追跡を再開できていることがわかる ((g), (h))。

このように、一度追跡対象を見失ったにも拘らず、正しい追跡が再開されたことを両者の追跡実験でそれぞれ確認できる。

以上のことから、環境属性は、追跡の安定化、特に追跡対象の初期化、再初期化に関して有用であると言える。

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、バルツェン窓 [2] 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。

文献

- [1] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society B*, Vol.39, pp.1-22, 1977.
- [2] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, pp.159-169, John Wiley & Sons, Inc., 2001.
- [3] F. van der Heijden, "Consistency Checks for Particle Filters," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.28, no.1, pp.140-145, 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] M. Isard and A. Blake, "CONDENSATION: Unifying low-



1300

1344

1376

1471

図 4.6: 追跡比較結果 (一様分布の利用)



1300

1344

1376

1471

図 4.7: 追跡比較結果 (環境属性の利用)



(a) # 315

(b) # 380

(c) # 734

(d) # 772

図 4.8: 追跡結果: 静的な遮蔽への対処



(e) # 564

(f) # 579

(g) # 601

(h) # 616

図 4.9: 追跡結果: 動的な遮蔽への対処

level and high-level tracking in a stochastic framework," Proc. European Conference on Computer Vision, vol.1, pp.893-908, 1998.

- [6] Y. Jin and F. Mokhtarian, "Data Fusion for Robust Head Tracking by Particles," Proc. IEEE International workshop on Visual Surveillance and Performance Evaluation and Tracking, pp.33-40, 2005.
- [7] Y. Kobayashi, D. Sugimura, Y. Sato, K. Hirasawa, N. Suzuki, H. Kage and A. Sugimoto, "3D Head Tracking using the Particle Filter with Cascaded Classifiers," Proc. the British Machine Vision Conference, pp.37-46, 2006.
- [8] 松元郁祐, 加藤丈和, 和田俊和, "Network Augmented Multi-sensor Association-CONDENSATION: CONDENSATIONの自然な拡張による3次元空間内での人物頭部の実時間追跡," 情報処理学会研究報告, 2005-CVIM-150-21, pp.161-168, 2005.
- [9] M. Sato and S. Ishii, "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 DII, vol.J88-DII, no.8, pp.1592-1600, 2005.
- [11] J. Vermaak, A. Doucet and P. Perez, "Maintaining Multi-Modality through Mixture Tracking," Proc. IEEE International Conference on Computer Vision, vol.2, pp.1110-1116, 2003.
- [12] C. Yang, R. Duraiswami and L. Davis, "Fast Multiple Object Tracking via a Hierarchical Particle Filter," Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.212-219, 2005.