逐次人物追跡のための事後確率を用いた人物抽出手法

Human Detection Using Posterior Probability
for Stochastic Human Tracking System

中西 航*・布施 孝志*

Wataru NAKANISHI and Takashi FUSE

Abstract : Recently sequential and stochastic human tracking method becomes popular in order to apply to complex situations that people are close to each other or are occluded by others. This paper proposes a human detection method for such tracking method, dealing with an appearance of human in observation area. We regard this human detection as an initial setting for stochastic tracking system. Firstly we detect human candidates and set initial distribution. Secondly we develop observation model using range and color information from sensors and calculate a posterior probability distribution. Thirdly we calculate expected value of human position and delete false positives, and then combine with tracking process. At last we apply the proposed method to the data acquired at the ticket gate of the railway station and show that the detection is succeeded under complex situations.

1. はじめに

精緻な施設設計や流動制御のために、多数の人物が相互に依存しながら行動するような複雑な状況における個々人の行動解析が望まれている。詳細な行動解析に不可欠な、複雑な状況における人物挙動データは、なお手作業により多大な時間や労力を要して取得されている現状である。そのため、取得の自動化、すなわち複数人物の自動追跡手法の確立が求められている。しかしながら、複雑な状況における人物の自動追跡は、従来よりチャレンジングな課題である。

一般的な動画像処理において、動物体追跡に対するアプローチは二つに大別される。各フレームで追跡対象となる物体を抽出しフレーム間で同一の物体を対応づける手法と、初期フレームで追跡対象となる物体を抽出し逐次追跡を行う手法である。いずれのアプローチにおいても対象物体の抽出手法が必要となる。近年の動物体追跡では、センシング技術の発達により取得できる観測データが膨大となり、一連の観測データ全てを同時に扱うことは困難である。ここでは、一度に

処理するデータ量の削減も重要となっている。そのため、各フレームにおける物体探索範囲を前フレームの情報に基づいて限定することが可能な逐次追跡手法が主流になりつつある。また、逐次追跡手法の利点として、対象物体が遮蔽されている状況や対象物体同士が接近している状況などの複雑な状況における高い追跡性能も挙げられる。

さらに近年では、逐次追跡の中でも確率的な手法が提案されている。確率的な手法においては、観測の不確実性や追跡対象物体の時系列遷移の非線形性を確率分布としてモデル化することにより、高い追跡精度を可能としている。例えば布施・中西 (2012) は、人物を追跡対象として、初期フレームにおける抽出を手動で行い、人物追跡を確率過程として扱う逐次追跡手法を構築した。この追跡手法は、後述する一般状態空間モデルの枠組みを利用し、非線形の確率分布を連立して構築しており、複雑な状況における高い追跡性能を実現している。しかし現在までのところ、このような確率過程の枠組みに人物抽出手法を組み込む研究は行われていない。

以上の背景より、本研究では逐次追跡手法における初期フレームを対象に、確率過程の枠組みで構築された人物追跡手法に統合可能な人物抽出手法を構築する。すなわち、観測領域に新たに出現する人物への対

*東京大学大学院工学系研究科社会基盤学専攻
「写真測量とリモートセンシング」VOL. 52, NO. 5, 2013

応としての抽出手法である。以下の章では、本研究のアプローチを説明し（2.），具体的なモデル化を行い（3.），実データに適用しその性能を確認する（4.）。

2. 本研究のアプローチ

2.1 人物追跡手法の概要

本節では、人物追跡手法（布施・中西，2012）の概要を述べる。この手法では、ステレオビデオカメラによる色・距離の観測情報を利用し、一般状態空間モデルの枠組みを用いて人物追跡を実現している。

一般状態空間モデルは、状態ベクトル $\mathbf{x}_t$ と観測ベクトル $\mathbf{z}_t$ から構成される。時刻 t における状態ベクトル $\mathbf{x}_t$ は直接の観測が不可能であるとする変数のベクトルであり、観測ベクトル $\mathbf{z}_t$ は観測情報を変数として持つベクトルである。ここで、状態ベクトルと観測ベクトルとの関係を表した観測モデル $p(\mathbf{z}_t|\mathbf{x}_t)$ と、状態ベクトルの時間推移を表したシステムモデル $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ とを、それぞれ条件付き分布としてモデル化する（樋口，2011）。そして、時刻 1 から t までの観測ベクトル $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t\}$ が得られた時に、時刻 t における状態ベクトル $\mathbf{x}_t$ を事後確率最大化基準により推定する。ここで、事後確率 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ は、ベイズの法則より、

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) \\ = p(\mathbf{z}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1} \quad (1)$$

と展開される。右辺のうち $p(\mathbf{z}_t|\mathbf{x}_t)$ は観測モデルを、 $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ はシステムモデルを、 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ は時刻 $t-1$ における推定の結果をそれぞれ示している。観測モデルでは、観測情報と状態ベクトルとのヒストグラム（色情報）および形状（距離情報）の類似度を比較する。また、システムモデルには歩行者挙動シミュレーションに用いられる離散選択モデルを導入している。

状態ベクトルと観測ベクトルは以後で扱うため、ここで具体的に示す。人物追跡手法では、状態ベクトルを追跡対象人物の全身として1楕円体で表現し（図1）、楕円体の中心座標 x, y, z および軸長 w, h, d の6変数で定義している。

$$\mathbf{x} = (x, y, z, w, h, d) \quad (2)$$

また、観測ベクトルはステレオビデオカメラから取得できる色・距離の情報であり、3次元座標 X, Y, Z お

図1 状態ベクトルと人物形状

よび色 r, g, b の6変数で定義している。画素 ij における観測ベクトルは、

$$\mathbf{z}_{ij} = (X_{ij}, Y_{ij}, Z_{ij}, r_{ij}, g_{ij}, b_{ij}) \quad (3)$$

となる。

一般に式(1)の計算には、初期条件として確率分布の初期分布を与える。この初期分布について次節で定式化を行う。

2.2 人物抽出問題の設定

既に述べたように、本研究は人物が出現した初期フレームにおける抽出を対象としている。これは、前節で述べた一般状態空間モデルに初期条件を与えることに相当する。ところで、式(1)に示した一般状態空間モデルの展開式は、初期条件として初期分布 $p(\mathbf{x}_0)$ が与えられ、時刻 1 から t までの観測ベクトル $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t\}$ が得られることを想定している。一方、本研究においては、 $t=0$ においても観測が存在するため、観測ベクトルとして $\mathbf{z}_{0:t} = \{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_t\}$ が得られる（図2）。従って、式(1)を得る時と同様に事後確率 $p(\mathbf{x}_t|\mathbf{z}_{0:t})$

図2 一般状態空間モデルにおける人物抽出

を展開した式において $t=1$ とすると、

$$\begin{aligned} p(\mathbf{x}_1|\mathbf{z}_{0:1}) &\propto p(\mathbf{z}_1|\mathbf{x}_1)p(\mathbf{x}_1|\mathbf{z}_0) \\ &= p(\mathbf{z}_1|\mathbf{x}_1) \int p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{z}_0)d\mathbf{x}_0 \end{aligned} \quad (4)$$

$$p(\mathbf{x}_0|\mathbf{z}_0) \propto p(\mathbf{z}_0|\mathbf{x}_0)p(\mathbf{x}_0) \quad (5)$$

の2式が得られる。ここで、本研究における人物抽出問題は式(5)として定式化される。すなわち、事後確率 $p(\mathbf{x}_0|\mathbf{z}_0)$ を得るために、初期分布 $p(\mathbf{x}_0)$ と $t=0$ における観測モデル $p(\mathbf{z}_0|\mathbf{x}_0)$ とを与える問題となる。事後確率を得ることにより、一般状態空間モデルで表される確率的な人物追跡システムに抽出手法を組み込むことが可能となる。

この際に、 $t=0$ における人物位置の推定値は、事後確率 $p(\mathbf{x}_0|\mathbf{z}_0)$ の期待値とする。また、抽出から追跡へ移行するためには、得られた事後確率を利用して式(4)を計算すればよい。なお、以降では区別のため、初期分布 $p(\mathbf{x}_0)$ の期待値を $\bar{\mathbf{x}}_{(prior)0}$ 、事後確率 $p(\mathbf{x}_0|\mathbf{z}_0)$ の期待値を $\bar{\mathbf{x}}_{(post)0}$ と表記する。同様に、 $\bar{\mathbf{x}}_{(post)t}$ は事後確率 $p(\mathbf{x}_t|\mathbf{z}_{0:t})$ の期待値である。

2.3 人物抽出の流れ

本研究における人物抽出の流れをフローチャートに示す(図3)。

まず、距離の観測情報に基づいて人物候補 $\mathbf{x}_0$ を抽出し、初期分布 $p(\mathbf{x}_0)$ を設定する(3.1)。ここで、一般的な確率過程では初期分布として正規分布を与えることが多い(樋口, 2011)ため、本研究でも同様とする。しかし、人物間の遮蔽や近接が起きる複雑な状況においては、人物候補 $\mathbf{x}_0$ の抽出精度が低下する。特に、位

置精度が低下する場合に、抽出された位置を平均値とする正規分布を初期分布として与えることは不適切である。提案手法では、この課題を解決するために、色・距離の観測情報に基づく観測モデル $p(\mathbf{z}_0|\mathbf{x}_0)$ を定義し、式(5)により事後確率 $p(\mathbf{x}_0|\mathbf{z}_0)$ を計算する(3.2)。これにより、人物候補の確率分布を正規分布からより観測情報に即した分布へと更新する。続いて、事後確率の期待値 $\bar{\mathbf{x}}_{(post)0}$ を計算し、誤抽出を消去する(3.3)。複雑な状況では誤抽出も生じやすいが、観測情報に即した事後確率を利用することにより精度の高い誤抽出消去が行える。最後に抽出結果を追跡プロセスと統合する(3.4)。以前のフレームで抽出され、既に追跡中の人物を再度抽出することのないように手法を構築する。各段階の詳細は次章で述べることにし、以下では既往研究との比較により本研究の特徴を述べる。

2.3.1 既往研究の整理

従来の人物抽出手法の多くは、色・距離情報を用いて人物候補 $\mathbf{x}_0$ の抽出のみを行うものである。多数の研究が蓄積されているため代表的な手法を概観する。いずれの手法においても、観測情報から人物と人物以外とを判別するための特徴量を定義している。ここでは、特徴量として点・線・面のいずれに着目するかによって整理する。

点を特徴量とする手法では、各画素が持つ画素値や各点が持つ距離の値をそのまま特徴量として利用する。色情報を利用する手法として、画素値を正規分布に従ってモデル化する PFinder (Wren *et al.*, 1997)、混合正規分布に従ってモデル化する MOG (Stauffer and Grimson, 1999)、背景画素値の変動幅に着目してモデル化する W4 (Haritaoglu *et al.*, 2000) がある。

図3 人物抽出の流れ

距離情報を利用する場合は、背景までの距離を既知とし、観測した点までの距離が背景までの距離より一定以上短ければ前景と判定する手法 (Harville *et al.*, 2001) が代表的である。いずれも、前景と判定された画素や点を統合し、最終的に人物を抽出する。

線の特徴量とする手法では、人物部位の輪郭形状に着目する。特に色情報に基づく手法は多数の研究がある (Dollar *et al.*, 2012)。代表例として、エッジ画像を作成し、局所領域におけるエッジ勾配の方向別強度に着目してヒストグラムを作成する HOG 特徴量を用いる手法 (Dalal and Triggs, 2005) や、局所領域内のエッジ画像の人物画像との共起性に着目した Shapelet 特徴量を用いる手法 (Sabzmejdani and Mori, 2007)、エッジ画像と上半身の輪郭形状との類似度に着目する手法 (Zhao and Nevatia, 2003) がある。また、距離情報から人物の足部の輪郭を抽出する手法 (中村ら, 2005) や上半身の形状を抽出する手法 (橋本ら, 2011) も存在する。

面の特徴量とする手法の代表例として Plan-View (Beymer, 2000; Darrell *et al.*, 2001) がある。これは、床平面を格子状に分割した 2 次元マップ (Plan-View Map と呼ぶ) を用いる手法である。このマップは、距離情報による 3 次元観測点を床平面に鉛直方向から投影して作成する。次元を 3 次元から 2 次元に圧縮することによる計算量削減のみならず、距離観測誤差の影響を軽減する効果が確認されている (Harville, 2004) ため、ステレオビデオカメラによる観測に対して広く使われている (Munoz-Salinas, 2008; 原口ら, 2010)。

2.3.2 本研究における人物抽出

本研究においても、ステレオビデオカメラから色・距離情報を取得する。そのため、前項の整理に基づき、人物候補 $\mathbf{x}_0$ の抽出手法として Plan-View を利用する。なお、点の特徴量とする手法では、前景の統合プロセスにおいて未抽出が起こりやすい (Zhao and Nevatia, 2004)。また、線の特徴量とする手法では、人物と判定すべきエッジ画像の事前学習が必要となるが、複雑な状況下で網羅的な学習を行うことは困難である (山下・藤吉, 2010)。これに対し、Plan-View を利用する積極的な理由として、画像上ではなく実空間上での人物候補の抽出を行えること、その結果として人物以外の物体を誤抽出する可能性が低いことが挙げ

られる。また、閾値の適切な設定を行えば人物の未抽出も少なくなる。

本研究の特徴は、得られた人物候補 $\mathbf{x}_0$ に対して、初期分布と観測モデルを定義することにより事後確率を算出する点にある。このことにより、前述の課題である、複雑な状況で抽出される人物候補の位置精度の低さや初期分布の実際の分布からのずれを、式(5)により修正できる手法となる。さらに、事後確率が分布として得られるため、確率過程において構築された追跡手法との統合に適した抽出手法となる。

3. 人物抽出手法のモデル化

本章では、提案手法の各段階についてモデルを構築する。以下では図 3 に従い、人物候補の抽出と初期分布の設定、観測モデルの定義と事後確率の計算、誤抽出の消去、追跡プロセスとの統合について、順に説明する。

3.1 人物候補抽出と初期分布設定

3.1.1 Plan-View Map

前章の通り、人物候補の抽出には Plan-View を利用する。ここで、説明のために図 4 に示す座標系を定義する。カメラの投影中心を床平面に鉛直方向から投影した点を原点とし、床平面を xz 平面に、鉛直上向きの軸を y 軸に設定する。また、 x 軸は 2 台のカメラの投影中心を床平面に鉛直方向から投影した点を通る直線とし、 z 軸は x 軸と直交するように定める。ステレオビデオカメラは基線がほぼ水平になるように設置されているため、カメラとこの座標系との回転行列は、カメラの床平面に対する撮影角度 ω と、カメラの設置高さ h_c が分かれば確定する。

床平面上の各格子に与える値によって 2 種類のマッ

図 4 座標系の設定

プを作成する。なお、Munoz-Salinas (2008) は Plan-View Map のみを用いて人物追跡を行うため色・距離情報から 3 種類のマップを作成しているが、本研究では抽出のみを対象とするため距離情報に基づく 2 種類のマップを作成し、色情報は観測モデルにおいて利用する(図 3)。また、同様の理由から、マップの平滑化を省略するなどマップ作成を簡略化している。

作成するマップは Occupancy Map と Height Map である。Occupancy Map は、投影される観測点数 O を各格子に与えたマップである。つまり、Occupancy Map は各格子から上方に伸びる直方体内部に存在する観測点数を表す。ただし、一つの観測点に対応する実際の物体の表面積は、カメラからの撮影距離の二乗に比例するため、投影時に以下の式(6)のように重み付けを行っている。Height Map は、投影される観測点の y 座標の各格子における最大値 H を示す。マップの作成にあたって投影対象とする点は、人物の大きさを考慮して床面からの高さが 2 m 未満にある点とする。以上より、格子 ij における O および H は、

$$\begin{aligned} O_{ij} &= \sum_{n \in ij} \{dist(n)\}^2 \\ H_{ij} &= \max_{n \in ij} \{y_n\} \end{aligned} \quad (6)$$

となる。ここで、 n は格子 ij 内に投影される点、 $dist(n)$ はカメラの投影中心から点 n までの距離、 y_n は点 n の y 座標である。

格子の大きさは、小さい値にすると観測した距離情報を高解像度のまま保持できる反面、観測誤差の影響が大きくなる。一方、大きい値にすると観測誤差の影響は小さくなるが、抽出結果の距離解像度が低下し位置精度が低下する。格子の大きさは、適用対象データに基づいて、試行実験により定める。

3.1.2 人物候補抽出と初期分布設定

前項で作成したマップに基づき、人物候補抽出を行う。Munoz-Salinas (2008) においては Occupancy Map と Height Map とを混合正規分布に基づいて評価するモデルとしているが、本研究では式(7)のように個別に評価する。Occupancy Map においては、 O の値が大きいほど人物候補の存在確率が高いと考えられるため、単調増加関数を設定する方が自然だからである。なお、本研究では、全ての観測点を投影してマップを作成する既往研究と異なり、前述の通りマップ作

成時に投影する観測点の範囲を床面からの高さで限定している。それぞれのマップの各格子における人物存在確率 $P_{occupancy}$ および P_{height} は、

$$\begin{aligned} P_{occupancy} &= \frac{1}{2} \left(1 + \operatorname{erf} \frac{O - \mu_o}{\sqrt{2\sigma_o^2}} \right) \\ P_{height} &= \exp \left(-\frac{(H - \mu_H)^2}{2 \times \sigma_H^2} \right) \end{aligned} \quad (7)$$

とする。ただし、

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (8)$$

は誤差関数である。また、 μ_o , σ_o , μ_H , σ_H は平均および分散を表すパラメータであり、適用対象データから得られる値から定める。

このモデルは、Occupancy Map については O が大きいほど人物存在確率が高いことを、Height Map については H が μ_H に近いほど人物存在確率が高いことを表現している。2 種類のマップから算出される人物存在確率の積

$$P_{new} = P_{occupancy} \cdot P_{height} \quad (9)$$

を各格子における人物存在確率とする。積をとることにより、床平面上の物体により $P_{occupancy}$ のみが大きい場合や、ステレオマッチング時のミスにより P_{height} のみが大きい場合における誤抽出を避けることができる。

全格子の中から P_{new} が大きい順に、格子を人物候補の位置として抽出する。本研究では、状態ベクトルは床平面である xz 平面に直立する楕円体として表現し、人物候補の床平面上の座標 (x_0, z_0) は格子の中心座標とする。これを、ある閾値 τ に対して、 $P_{new} > \tau$ を満たさなくなるまで繰り返す。この際に、ある格子が人物候補として抽出された場合には、人物の平面的な大きさを考慮して、抽出された格子の中心座標から 50cm 未満に中心座標がある格子についてはその時点で $P_{new} = 0$ とする。閾値 τ の値は大きくすると誤抽出が減少するが未抽出が増加し、小さくすると未抽出が防げるが誤抽出が増加する。本研究では、後述の誤抽出消去プロセスを有するため、未抽出が生じないように定める。

また、Plan-View からの人物候補 $\mathbf{x}_0$ の抽出にあたっては、状態ベクトル $\mathbf{x}$ に含まれる 6 変数のうち、床平面上の位置を表す変数 x と z のみを対象とする。残りの変数である、高さ方向の位置を表す変数 y 、楕円

体の大きさを表す変数 w, h, d については一律の値を与えることとする。具体的には、人物候補の床平面座標 x_0 および z_0 に、 $y_0=1.0$ [m], $w_0=0.4$ [m], $h_0=1.6$ [m], $d_0=0.3$ [m] を与え、人物候補 $\mathbf{x}_0$ とする。

ここに、初期分布として正規分布を仮定し、 $p(\mathbf{x}_0)$ を生成する。変数間の共分散は無いものとし、 $p(\mathbf{x}_0) \sim N(\bar{\mathbf{x}}_{\text{prior}0}, \boldsymbol{\sigma}_{\text{init}}^2)$ とした。分散の値は試行実験に基づいて決定する。また、人物の出現を対象とするため、人物抽出は観測領域の周縁部においてのみ行う。以下では、抽出対象とする領域を抽出領域と呼ぶ。

3.2 観測モデルと事後確率の計算

本節では、前節で得られた $p(\mathbf{x}_0)$ を観測モデルによって修正し、事後確率分布を得る。観測モデルは、色情報に基づく観測モデル $P_{\text{color}}(\mathbf{z}_0|\mathbf{x}_0)$ と距離情報に基づく観測モデル $P_{\text{range}}(\mathbf{z}_0|\mathbf{x}_0)$ の両方をモデル化し、その積

$$P(\mathbf{z}_0|\mathbf{x}_0) = P_{\text{color}}(\mathbf{z}_0|\mathbf{x}_0) \cdot P_{\text{range}}(\mathbf{z}_0|\mathbf{x}_0) \quad (10)$$

により定義する。

3.2.1 色情報に基づく観測モデル

色情報に基づく観測モデルはカラーヒストグラムの類似度に従う確率分布とし、人物候補と背景との類似度が低いほど高い値を返すモデルとする。背景情報は、Wren *et al.* (1997) に基づき、背景の画素値をモデル化して作成する。画素値 r, g, b の3色それぞれについて観測情報として画素値 C が得られるたびに、画素値の平均 μ と分散 σ^2 を、背景である確率 P_{bg} と更新率 α に基づいて更新する。具体的には、時刻 t において、

$$\begin{aligned} \mu_t &= \alpha_t C_t + (1 - \alpha_t) \mu_{t-1} \\ \sigma_t^2 &= \alpha_t (C_t - \mu_t)^2 + (1 - \alpha_t) \sigma_{t-1}^2 \\ P_{bg,t} &= \exp\left(-\frac{(C_t - \mu_t)^2}{2 \times \sigma_t^2}\right) \\ \alpha_t &= 0.1 \times P_{bg,t} \end{aligned} \quad (11)$$

とする。更新率 α は、新しい観測に基づく画素値と背景である確率が高い画素値とを重み付けする変数である。

人物候補のヒストグラム d_0 は、楕円体 $\mathbf{x}_0$ 内部の画素を対象に作成する。また、背景のヒストグラム d_{bg} は、楕円体 $\mathbf{x}_0$ 内部の画素が画像上に投影される点について、式(11)の第1式を用いて作成する。2個のヒスト

グラムをそれぞれ正規化した後、ヒストグラム間の非類似度を示す指標である Bhattacharyya 距離

$$B = 1 - \sum_m \sqrt{d_{0,m} d_{bg,m}} \quad (12)$$

を求める。ここで、 m は取り得る画素値、 $d_{i,m}$ はヒストグラム d_i における画素値 m の相対頻度を示す。 r, g, b の3色について式(12)を計算し、それらの積を色情報からの観測モデル $P_{\text{color}}(\mathbf{z}_0|\mathbf{x}_0)$ とする。

3.2.2 距離情報に基づく観測モデル

距離情報からの観測モデルは、布施・中西 (2012) の追跡手法と同一とする。楕円体の画像平面への投影面に含まれる画素の3次元座標と、楕円体表面の3次元座標との比較によってモデル化する。いま、対象となる画素の3次元座標上の地点を P とした時、楕円体中心 O と P との距離を $d(P)$ とする。また、 O を始点に点 P を通る半直線が楕円体表面と交わる点を P' とした時、 O と P' との距離を $\hat{d}(P)$ とする。この時、 $d(P)$ と $\hat{d}(P)$ との差分の和が小さいほど抽出対象の存在確率が高いことから、

$$P_{\text{range}}(\mathbf{z}_0|\mathbf{x}_0) = 1 - \left(\frac{1}{I} \sum_P (d(P) - \hat{d}(P))^2 \right) \quad (13)$$

とする。ここで、 I は対象となる画素数である。なお、 $P_{\text{range}}(\mathbf{z}_0|\mathbf{x}_0)$ が負にならないようにするため、 $|d(P) - \hat{d}(P)| > 1$ となる画素は、 $|d(P) - \hat{d}(P)| = 1$ として計算する。

3.2.3 事後確率分布と人物位置の期待値の算出

前節で得られた初期分布 $p(\mathbf{x}_0)$ と観測モデル $p(\mathbf{z}_0|\mathbf{x}_0)$ (式(10)) とから式(5)の計算を行い、事後確率 $p(\mathbf{x}_0|\mathbf{z}_0)$ を算出する。この計算は非線形の確率分布を含むため、パーティクルフィルタ (Gordon *et al.*, 1993; Israd and Blake, 1998) を利用して行う。また、事後確率分布を得ると同時に、事後確率の期待値 $\bar{\mathbf{x}}_{\text{post}0}$ を算出する。次節ではこの算出結果を利用して誤抽出の消去を行う。

3.3 抽出時・追跡の初期段階における誤抽出消去

本節では、以上の手順で抽出された人物候補から、誤抽出された候補を消去する。事後確率の期待値である $\bar{\mathbf{x}}_{\text{post}0}$ の推定値から人物候補を表す楕円体中心間の床平面上の距離を算出し、50cm 未満となる場合には同一人物であるとみなす。そして、最大の P_{new} から

生成された人物候補のみを残し、他の候補を消去する。また、この時、図3に示すように追跡プロセスから追跡中の人物位置リストを受け取っているため、人物候補と人物位置リスト上にある人物との距離が50cm未満となった場合には、追跡中の人物のみを残し、人物候補は誤抽出として消去する。

追跡の初期段階においても同様の処理を行う。すなわち、事後確率の期待値である $\bar{\mathbf{x}}_{(post)t}$ を算出するたびに、追跡の初期段階にある全ての人物候補を対象に楕円体中心の床平面上の距離を算出し、50cm未満となる場合には同一人物であるとみなすとともに、誤抽出と判定された人物候補は消去する。本研究では、抽出された人物候補が抽出領域内において追跡されている間を追跡の初期段階として扱う。そして、追跡の初期段階が終了した時点でも追跡が継続されている人物候補を、最終的に抽出された人物として確定する。

3.4 追跡プロセスとの統合

構築した抽出プロセスを追跡プロセスに統合するため、人物候補抽出にあたって既に追跡中の人物を抽出しないように処理を行う。既に追跡中の人物の中心座標から床平面上で50cm未満に中心座標がある格子を抽出対象外領域とし、あらかじめ $P_{new}=0$ とする。

また、式(4)の計算にあたって、システムモデル $P(\mathbf{x}_1|\mathbf{x}_0)$ にはランダムウォークを仮定する。これは、2.1で述べた追跡手法（布施・中西，2012）におけるシステムモデルが人物の現在速度を必要とすることに対し、 $t=0$ の時点では抽出された人物の速度が不明だからである。なお、 $t=1$ において $\bar{\mathbf{x}}_{(post)1}$ の推定値と $\bar{\mathbf{x}}_{(post)0}$ の推定値との差分が人物の現在速度として得られるため、 $t=1$ 以降は追跡手法と同様のシステムモデルを用いる。

4. 適用

4.1 撮影条件

構築した提案手法を、ステレオビデオカメラで撮影した実動画像へ適用する。撮影は、2010年11月29日の朝ラッシュ時間帯に、東急田園都市線たまプラーザ駅改札で行った。本研究で使用したステレオビデオカメラは、2台の SONY-DFW (120万画素) を約1mの間隔で設置し、事前にキャリブレーションを行って構成したものである(表1)。このカメラを用いて、斜めに見下ろすような撮影を行った(図5)。カメラの床面に対する撮影角度と設置高さを、床面6点までの距離情報を利用して、 $\omega=0.63$ [rad] および $h_c=12.0$ [m] とそれぞれ算出した。

4.2 マップ作成とパラメータ設定

撮影データに対し、Occupancy Map および Height Map を作成する。床平面が既知であるため、前述の通り、図4の座標系で $0 < y < 2$ を満たす観測点のみを対象とした。Plan-View Map の格子の大きさは、3, 5, 10 [cm] の3通りを試し、観測誤差の影響と抽出精度とのバランスから5cmとした。その上で、各マップの持つ値と実際の人物位置との関係を確認し、式(7)におけるパラメータを $\mu_0=7500$, $\sigma_0=1000$, $\mu_H=1.5$, $\sigma_H=0.5$ とした。作成されたマップの例を、一部分を拡大して図6に示す。Occupancy Map は黒色が濃いほど O が大きいこと、Height Map は黒色が濃いほど H が大きいことを示す。概ね人物位置の床面上が濃くなっていることが確認できる。人物抽出に際する閾値 τ は、前述の通り未抽出が生じないように定める。 $\tau=0.2, 0.25, 0.3$ の3通りを試行した結果から、 $\tau=0.2$

表1 ステレオビデオカメラのキャリブレーション結果

	カメラ1 (基準カメラ)	カメラ2
焦点距離 f [mm]	8.292494	8.327826
1次レンズ歪み係数	0.001402	0.001274
2次レンズ歪み係数 [mm ⁻²]	-3.4×10^{-5}	-3.0×10^{-5}
回転行列	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.999992 & 0.000853 & 0.004079 \\ -0.00087 & 0.999996 & 0.002758 \\ -0.00408 & -0.00276 & 0.999988 \end{bmatrix}$
並進ベクトル [mm]	$[0 \ 0 \ 0]^T$	$[-999.779 \ -1.19145 \ 6.682257]^T$
画像中心座標 [pix]	(666.5752, 486.8504)	(694.4858, 509.4942)

図5 取得画像例

図7 抽出範囲の設定

画像の一部

左画像と同範囲の Occupancy Map

左画像と同範囲の Height Map

図6 Occupancy Map および Height Map

図8 適用結果

とした。

初期分布 $p(\mathbf{x}_0)$ の分散および初期のシステムモデル $P(\mathbf{x}_1|\mathbf{x}_0)$ の分散は、理論的に定めることはできないため、試行実験により決定した。具体的には、まず、初期分布を $\sigma_{init}=(5, 25, 5, 1, 2.5, 1)$ [cm] とした。この際に、位置については、Plan-View からの抽出結果に基づいている x, z に比較的小さい値を、本来個人差があるものの一律の値を与えている y に比較的大きい値を想定した。また、形状 w, h, d については、接近している別の人物が楕円体内に含まないようにするため小さい値を想定した。次に、初期のシステムモデルについて $\sigma_{sys}=(10, 2.5, 10, 0.5, 0.5, 0.5)$ [cm] とした。この際に、 x, z は人物が 1 フレーム間に進む距離を想定した。また、 y, w, h, d は短時間での変化は微小であることから小さい値を想定した。

4.3 適用結果と考察

提案手法を、120秒間（900フレーム）を対象として適用した。抽出領域は観測領域の外縁部とし、図7に黄色で示している。観測領域の外側から抽出領域に進入してきた人物171人を抽出対象とした。抽出領域への進入場所の内訳は、北口が56人、南口が55人、1番線が37人、2番線が23人である（表2）。

適用結果の一部を図8に示す。図中の楕円は、抽出された人物候補または該当フレームにおいて追跡中の人物に対応する状態ベクトル $\bar{\mathbf{x}}_{(post)t}$ の楕円体を画像に投影した領域の外縁を示している。青色の楕円は、それぞれのフレームにおいて誤抽出されている人物候補を示す。これらの多くは、後述するように追跡の初期段階において消去することに成功している。

提案手法による人物抽出性能の検証は目視で行った。推定された抽出結果の楕円体と実際の人物領域とが、画像上で1画素以上重なっている場合に抽出成功

とした。1画素以上とした理由は、カメラから遠い位置の人物や、一部が遮蔽されている人物では、当該人物に含まれる可能性のある画素が10画素程度まで少なくなる場合があるためである。ただし、抽出結果の楕円体が抽出対象人物と重なっている画素数よりも、抽出対象以外の人物と重なっている画素数の方が多い場合は、失敗とした。

検証の結果、171人全てが抽出領域内において一度は人物候補として抽出されており、閾値 τ の設定が妥当であることを確認した。また、抽出領域内において1人の人物に対して複数の人物候補が抽出されたケースは64回あった。しかし、これらの人物候補のうち、追跡の初期段階が終了する時点でも消去されず誤抽出となった人物は11人とどまっている。反対に、誤抽出ではないにも関わらず抽出から追跡へ移行する際に消去され、未抽出となった人物が5人存在する。以上から、抽出結果は171人中155人となった（表2）。また、事後確率を利用する効果を確認するため、初期分布のみを用いた抽出との比較を行った。3.1.2で述べたとおり、初期分布のみを用いる手法では、 τ の値を変更しても未抽出と誤抽出を同時に減らすことはできない。提案手法と同じ未抽出人数（5人）を実現する τ では、誤抽出が61人となった。また、提案手法と同じ誤抽出人数（11人）を実現する τ では未抽出が23人となった。この比較結果からも、事後確率を用いる提案手法が有効であるといえる。

特に、図8の各フレームの右上側に示すような混雑した状況においても、未抽出と誤抽出の両方を回避することに成功している。具体的には、frame #745において、緑色の線で囲まれた領域から8個の人物候補が抽出された。この領域に実際に存在する人物は6人である。つまり、全員の抽出に成功し、2個が誤抽出である。判読性から図8には表示していないが、抽出された人物候補には固有の番号を付与している。この番号により確認した結果、これら2個の誤抽出は frame #750の時点で消去されていた。また、frame #750で発生した2個の誤抽出には frame #745における誤抽出とは別の番号が付与されており、これらも以降で同様に消去されることを確認した。

なお、全フレームを通した全体的な傾向として、誤抽出の多くは、目視では何も存在しない位置に観測ベクトルの3次元座標が存在する時、すなわち3次元座標を求めるステレオマッチングが誤っている時に発生

表2 抽出結果

進入場所	対象人数	抽出成功人数	誤抽出人数	未抽出人数
北口	56	52	4	0
南口	55	52	1	2
1番線	37	33	2	2
2番線	23	18	4	1
合計	171	155	11	5

していた。誤った3次元座標が式(7)における $H=\mu_H$ の付近に複数存在すると、 P_{height} と $P_{occupancy}$ がともに比較的大きな値となる。その結果、 P_{new} が閾値 τ よりも大きくなり誤抽出となる。このような Plan-View を利用しても対応できないステレオマッチングのミスに対しては、より正確なステレオマッチング点のみを用いて人物抽出を行うことが考えられる。また、未抽出となった場合の多くは、人物候補の抽出時に反対方向に向かって進んでいる追跡中の人物が接近している状況にあった。そのため、追跡への移行時に、抽出した人物候補と接近していた人物とを誤認している。人物の抽出位置に応じて $\tau=0$ におけるシステムモデルに指向性を持たせることによる対応が考えられる。

5. おわりに

本研究では、確率過程の枠組みで構築された人物逐次追跡手法において、観測領域に出現する人物抽出を確率過程の初期設定問題と位置づけ、人物抽出手法を構築した。まず、人物候補を抽出し初期分布を設定した上で、複数観測情報からの観測モデルにより事後確率分布を求めた。これにより、初期分布と比較して、より観測情報に即した確率分布を用いることができる。その結果、誤抽出消去が可能となり、初期分布のみを用いる手法では困難であった未抽出と誤抽出との同時削減を可能とした。最終的には、抽出プロセスと追跡プロセスを確率過程において統合し、提案手法を構築した。さらに、提案手法を実データに適用し、複雑な状況における有効性を確認した。この提案手法により、長時間にわたる複数人物の自動抽出・追跡が実現可能となる。なお、提案手法内で利用しているモデルは観測条件に応じて組み替えられるため、観測可能な情報に基づいて人物候補の抽出手法や観測モデルのモデル化を行うことができる特徴を有する。

今後の課題として、さらなる抽出精度の向上が挙げられる。誤抽出を削減するためには、Plan-View におけるマップ作成手法や人物存在確率のモデル改良が求められる。これに対しては、ステレオマッチングのマッチングミスの影響を軽減するためにマッチング精度の良い点のみを用いてマップを作成することや、マップの各セルにおける人物存在確率を近傍のセルの値も考慮して計算するモデルに改良することが考えられる。また、未抽出を削減するためには、追跡の初期段階に

におけるシステムモデルの検討が必要である。例えば、現在はランダムウォークとしているが、抽出場所から観測領域の中央方向に向かうものに変更することが考えられる。さらに、本研究とは異なる対象への適用では、対象場の人物密度や速度等の状況に応じたモデル化やパラメータ設定が必要となる。この設定にかかる労力を軽減するためには、多様な状況にある対象への適用を通して、場の状況とモデルやパラメータとの関係を把握し体系化していくことが望まれる。

謝辞：研究に際し清水英範教授（東京大学）より貴重なご意見を頂いた。城石典明氏（㈱東急総合研究所）、横内稔充氏（東京急行電鉄㈱）、山口博義氏（コマツエンジニアリング㈱）には現場撮影でご協力いただいた。また、本研究は、科学研究費(若手研究(B)：22760401)の助成を受けたものである。ここに記して感謝の意を表する。

(受付日2013.5.1, 受理日2013.9.13)

参考文献

- 中村克行, 趙 卉菁, 柴崎亮介, 坂本圭司, 大鋸朋生, 鈴川尚毅, 2005. 複数のレンジスキャナを用いた歩行者トラッキングとその信頼性評価, 電子情報通信学会論文誌, J88-D-II (7), pp.1143-1152.
- 橋本 潔, 加賀屋智之, 片岡裕雄, 里 雄二, 田藤雅基, 大島京子, 藤田光子, 青木義満, 2011. 姿勢変動を考慮した基幹部位追跡と肩位置推定, 画像の認識・理解シンポジウム (MIRU2011) 論文集, pp. 668-674.
- 原口真悟, 福司謙一郎, 熊澤逸夫, 2010. 多数のステレオカメラと色情報を用いた人物追跡手法, 電子情報通信学会技術研究報告 (ヒューマン情報処理), 109(471), pp.229-234.
- 樋口知之之編著, 2011. データ同化入門—次世代のシミュレーション技術—, 朝倉書店.
- 布施孝志, 中西 航, 2012. 歩行者挙動モデルを統合した人物自動追跡手法の構築, 土木学会論文集 D3, 68(2), pp.92-104.
- 山下隆義, 藤吉弘亘, 2010. Soft Decision Feature による姿勢変化に対応した人物追跡, 情報処理学会論文誌 CVIM, 51(9), pp.1940-1950.
- Beymer, D., 2000. Person Counting Using Stereo, *Proc. of IEEE Workshop on Human Motion*, pp.

- 127-133.
- Dalal, N. and Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, pp.886-893.
- Darrell, T., Demirdjian, D., Checka, N. and Felzenszwalb, P., 2001. Plan-view trajectory estimation with dense stereo background models, *Proc. of IEEE International Conference on Computer Vision*, 2, pp.628-635.
- Dollar, P., Wojek, C., Schiele, B. and Perona, P., 2012. Pedestrian Detection: An Evaluation of the State of the Art, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4), pp.743-761.
- Gordon, N.J., Salmond, D.J. and Smith, A.F.M., 1993. Novel approach to nonlinear / non-Gaussian Bayesian state estimation, *Radar and Signal Processing*, IEE Proceedings F, 140(2), pp.107-113.
- Haritaoglu, I., Harwood, D. and Davis, L.S., 2000. W4: real-time surveillance of people and their activities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), pp.809-830.
- Harville, M., Gordon, G. and Woodfill, J., 2001. Foreground Segmentation Using Adaptive Mixture Models in Color and Depth, *Proc. of IEEE Workshop on Detection and Recognition of Events in Video*, pp.3-12.
- Harville, M., 2004. Stereo person tracking with adaptive plan-view templates of height and occupancy statistics, *Image and Vision Computing*, 22, pp.127-142.
- Israd, M. and Blake, A., 1993. Condensation — Conditional density propagation for visual tracking, *International Journal of Computer Vision*, 29(1), pp.5-28.
- Munoz-Salinas, R., 2008. A Bayesian plan-view map based approach for multiple-person detection and tracking, *Pattern Recognition*, 41(12), pp.3665-3676.
- Sabzmeydani, P. and Mori, G., 2007. Detecting Pedestrians by Learning Shapelet Features, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8.
- Stauffer, C. and Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, pp.246-252.
- Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A., 1997. Pfunder: real-time tracking of the human body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp.780-785.
- Zhao, T. and Nevatia, R., 2003. Bayesian Human Segmentation in Crowded Situations, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, pp.459-466.
- Zhao, T. and Nevatia, R., 2004. Tracking multiple humans in complex situations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), pp.1208-1221.