

複数対象トラッキングと複数ダイナミクスの同時推定法

石黒 勝彦[†] 山田 武士[†] 上田 修功[†]

[†] NTT コミュニケーション科学基礎研究所, 京都府

E-mail: †{ishiguro,yamada,ueda}@cslab.kecl.ntt.co.jp

あらまし 従来の複数対象トラッキング手法は、全ての追跡対象について一つのダイナミクスモデルを適用することが多い。しかし、シーン内に存在する全ての対象が常に同一のダイナミクスに従うとは限らない。この問題に対処するためには複数のダイナミクスパターンが必要となるが、シーンの解析前に適切な数のダイナミクスパターンを全て人手で決定する事は困難であり、自動的に学習できることが望ましい。本論文では、複数の移動対象をトラッキングすると共に、各対象のダイナミクスパターンをクラスタリングして学習する確率的な生成モデルを提案する。人工データ、および実動画データを用いた実験を通じて、提案モデルがトラッキングとダイナミクスの学習を同時に実現可能であること、またこのことでトラッキング自体の性能も向上することを示した。

キーワード 複数対象トラッキング, クラスタリング, 生成モデル, ノンパラメトリックベイズ, モデル推定

Multi-target tracking with multiple dynamics estimation

Katsuhiko ISHIGURO[†], Takeshi YAMADA[†], and Naonori UEDA[†]

[†] NTT Communication Science Laboratories, Kyoto, 619-0237, Japan

E-mail: †{ishiguro,yamada,ueda}@cslab.kecl.ntt.co.jp

Abstract We present an on-line probabilistic generative model that simultaneously deals with both the clustering and the tracking of an unknown number of moving objects. We assume that time series data are composed of a i) time-varying number of objects and that ii) each object is governed by an unknown number of different patterns of dynamics. We utilize a nonparametric Bayesian model for clustering patterns of dynamics, and this clustering in turn improves the tracking. In experiments, proposed model successfully learned the hidden patterns and obtained better tracking results than conventional models without clustering.

Key words Multi-target tracking, Clustering, Generative models, Nonparametric Bayes, Model estimation

1. ま え が き

複数移動対象の追跡はコンピュータービジョンにおいて重要な問題の一つであり、監視カメラ画像の解析や動画コンテンツの検索等応用のために多くの研究がなされている [1]~[4].

複数対象のトラッキング手法としては、不連続な観測データ系列に状態空間モデルのダイナミクスを当てはめることで、連続な (フィルタリングされた) 対象軌跡を確率的に推定する手法が一般的である [5], [6]. 通常、当てはめに用いられる状態空間モデルは既知とされて運用される。一方で、追跡対象の数は単一ながらフィッティングすべきダイナミクスを推定しながらフィルタリングを行う手法も提案されている [7], [8]. この手法では、ダイナミクスの混合を許容した上で混合数と各ダイナミクスのパラメータを推定できる。

これら既存手法をさらに拡張したモデルとして、混合ダイナミクスを推定しながら複数対象のトラッキングを実現する手法

を考えることができる。そのような手法が開発できれば、i) 多くの移動物体がシーン中で観察され ii) 対象ごとに異なるダイナミクスに従って行動するような、より広範な動画像データでの追跡をモデル化することが可能となり、対象軌跡の予測、ひいては時系列データの理解に有利に働くと思われる。

本論文では、ダイナミクスパターンをクラスタリングしてその数とパラメータを推定すると同時に未知数の対象に対するトラッキングも行うことができる、確率的な生成モデルを提案する [9]. 提案モデルでは、確率的な複数対象トラッキングモデルに、Dirichlet Process Mixture (DPM) によるクラスタリングモデルを採り入れる。この組み合わせによって、ダイナミクスパターンの数とそのパラメータを推定しつつ、同時に個々の移動対象軌跡を推定することが可能となる (図 1.)。パターンはクラスタリングによって決定され、各要素は異なるパラメータを持ったカルマンフィルタである。隠れ変数と追跡対象の状態量の推定にはパーティクルフィルタ [10], [11] を用いる。人工

データおよび実動画像データを用いた実験を通じて、提案モデルの効果を確認する。

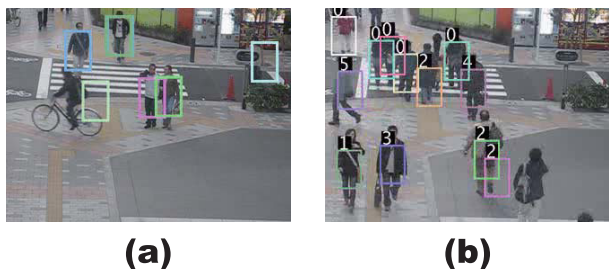


図1 既存のトラッキングシステムの結果(a)と提案モデルの結果(b). 矩形上部の数字は対象の現在のダイナミクスパターンインデックスを表す. 提案モデルでは複数対象の追跡と同時に, 各対象の運動パターンを逐次的に推定する.

2. 先行研究

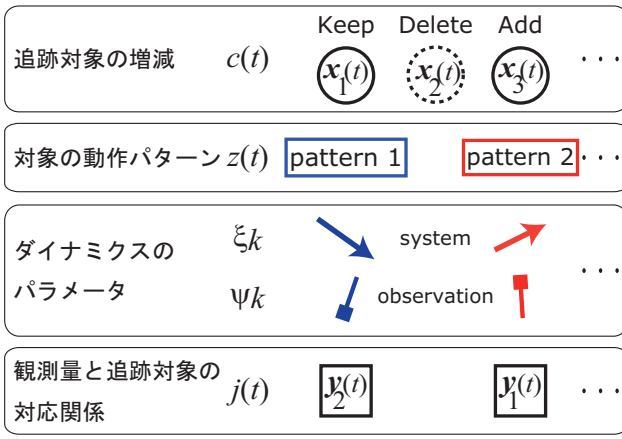
最初に, Särkkä らによる複数対象トラッキングのモデル [6] を説明する. このモデルでは, 時間とともに追跡対象の数が増減する状況下での複数対象トラッキングを実現する例である.

各時刻で複数の追跡対象および観測が存在するため, 隠れ状態には i , 観測には m の添字をつけて区別する. i 番目の対象 (隠れ状態) は $x_i(t)$, m 番目の観測を $y_m(t)$ とする. 複数の追跡対象と観測が存在する場合, どの対象がどの観測を出力したのかという対応関係 (data Association) を解決しなければならない. このモデルでは, 追跡対象の数が増減する状況下での data association を解決するため, Rao-Blackwellized Particle filter を用いた確率モデルを提案している. ただし, 彼らの研究においては, 全ての追跡対象についてその状態空間モデルが事前に固定された一つの Kalman filter [11] で表現できるものと仮定されている. 先に述べたとおり, この仮定 (制約) は一般に正しいとは限らない.

次に, Caron らによるモデル [8] を説明する. このモデルは, 単一の追跡対象のみが存在する, あるいは single stream の時系列データのフィルタリングを念頭においたものである. 対象の隠れ状態は, 未知の混合ダイナミクスパターンから生成されたものと仮定される. ここで未知なのはダイナミクスの混合数ならびに各ダイナミクスのパラメータであり, 隠れ状態の推定とともにこれらの要素も同時に推定することを目的とする.

Särkkä らのモデルと対照的に, 隠れ状態や観測は単一であるため $x(t)$ や $y(t)$ には添字が必要ない. 一方でダイナミクスが複数存在するため, 添字 k をつけて区別する. 各ダイナミクスは状態空間モデル, 具体的には Kalman filter と仮定する. 状態モデルのパラメータ (正規分布の平均ベクトルおよび共分散行列) を $\{\xi_k\}$, 観測モデルのパラメータ (やはり平均ベクトルと共分散行列) を $\{\psi_k\}$ で表す. $x(t)$ ($y(t)$) はパラメータ ξ_k (ψ_k) で特徴づけられた分布に従って生成される. データ生成過程を簡単に説明すると, 各時刻 t において, まずダイナミクスパターンインデックスを表す隠れ変数 $z_t = k$ が生成される. そして, z_t に対応したパラメータ ξ_k , ψ_k がサンプリ

隠れ変数を生成



隠れ状態と観測を生成

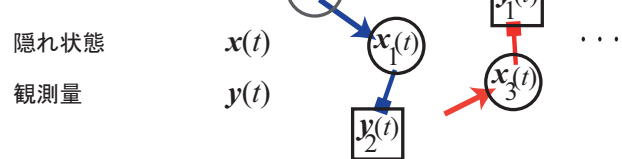


図2 提案するデータ生成モデルの概念図. 複数対象の増減やダイナミクスの情報を表現する隠れ変数, および隠れ状態と観測を示す

ングされ, これらのパラメータのもとで隠れ状態と観測が生成される.

このモデルの特徴は, ダイナミクスの混合分布を表現するために Dirichlet Process Mixture (DPM) を用いた点である. DPM はノンパラメトリックベイズモデルの一つで, 各混合コンポーネントの性質 (パラメータ) の推定だけでなく, 混合数についてもベイズ的枠組で同時に推定できるというメリットがある. 従って, このシステムは事前に混合モデルの混合数や各コンポーネントのパラメータを指定することなくダイナミクスのパターンを学習可能である. 著者らは論文中で MCMC を用いたオフライン推定と Particle filter を用いたオンライン推定の2つの推論方法を示している. しかし, 本モデルには data association を解決する仕組みがないために複数の追跡対象が同時に存在する時系列データには適用できない.

3. 提案法

続いて, 提案するモデルとその推定方法を説明する. 本稿では大まかな説明にとどめるが, モデルの詳細については [9] を参照のこと.

3.1 モデル

我々の提案するモデル (図2) は, 隠れ変数とデータ (隠れ状態, 観測) およびパラメータからなる生成モデルである. i 番目の対象の隠れ状態は $x_i(t)$, m 番目の観測は $y_m(t)$ とする. 各時刻 t において, パラメータから隠れ変数とデータを生成するプロセスを説明する. まず最初に追跡対象数の増減 (生成と消滅) を表現する隠れ変数 $c_i(t)$ を生成する (式 (4)). $c_i(t) = 1$ の場合 i 番目の追跡対象がシーン中に存在し $c_i(t) = 0$ の場合

シーン中に存在しないものとする。この変数 $c(t)$ を制御することで、時系列データ内に存在する追跡対象の軌跡数を時間ごとに変化させることができる。

続いて、シーン中に存在する追跡対象 ($c(t)$ で決定された) がこの時刻で従うダイナミクスを決定する。すなわち、時刻 t において i 番目の追跡対象 $\mathbf{x}_i(t)$ を支配するダイナミクスパターン of インデックス $z_i(t) = k$ の値を確率的に決定する (式 (6))。このプロセスは DPM の実現例の一つである Chinese Restaurant Process (CRP) [12] を用いてモデル化する。

次に、実際のダイナミクスパラメータ ξ_k と ψ_k をサンプリングする (式 (7))。本モデルでは状態空間モデルとして Kalman filter を用いる。 ξ_k と ψ_k はそれぞれ状態モデルおよび観測モデルの正規分布パラメータ、すなわち平均ベクトルと共分散行列である。

隠れ状態 $\mathbf{x}_i(t)$ は、対象 i がシーン中に存在する場合のみ生成される。インデックス $z_i(t) = k$ で決定したパラメータ ξ_k に従って隠れ状態の時間発展を計算する。

続いて m 番目の観測量 $\mathbf{y}_m(t)$ の生成のために、data association 変数 $j_m(t)$ (式 (5)) を生成する。 $j_m(t)$ は [6] で導入された $\mathbf{x}_i(t)$ と $\mathbf{y}_m(t)$ の間の data association を表現する隠れ変数である。 m 番目の観測量が i 番目の隠れ状態に対応する場合 $j_m(t) = i$ と表す。この時 $\mathbf{y}_m(t)$ は $\mathbf{x}_i(t)$ とパラメータ $\psi_{z_i(t)}$ で規定された観測モデル方程式から生成する。このプロセスを全ての t について繰り返すことで、複数の異なるダイナミクスで支配された、複数対象の行動軌跡の重ね合わせからなる、一般の複数対象の行動時系列データを生成することが可能である。

3.2 Particle Filter を用いた隠れ変数の推定手法

続いて複数対象のトラッキングとダイナミクスモデルのクラスタリング・パラメータの推定をオンラインで行う方法を説明する。これ以降、各シンボルの大文字でもって、時刻 1 から時刻 t までの集合を表す。例えば、 $\mathbf{X}(t) = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)\}$ であり、 $\mathbf{Y}(t) = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(t)\}$ とする。

目的は事後分布 $p(\mathbf{x}(t)|\mathbf{Y}_t)$ の推定である。時刻 t での隠れ変数集合を $\phi(t) = \{c_i(t), \{j_m(t)\}, \{z_i(t)\}, \{\xi_k(t)\}, \{\psi_k(t)\}\}$ と書く事として、 $\Phi(t) = \{\phi(1), \phi(2), \dots, \phi(t)\}$ とする。この記号を用いて、事後分布を次のように Particle filter で近似する:

$$p(\mathbf{x}(t)|\mathbf{Y}(t)) = \int p(\mathbf{x}(t), \Phi(t)|\mathbf{Y}(t)) p(\Phi(t)|\mathbf{Y}(t)) d\Phi(t) \\ \triangleq \sum_{s=1}^S p(\mathbf{x}(t), \Phi(t)^{(s)}|\mathbf{Y}(t)) w(t)^{(s)} \quad (1)$$

ここで S はパーティクルの総数である。本モデルでは先行研究 [5], [6] に従って、状態空間モデルを Kalman filter で定義するので、 $p(\mathbf{x}(t), \Phi(t)|\mathbf{Y}(t))$ は簡単に計算可能である。この分布をサンプリングされた隠れ変数 $\Phi(t)^{(s)}$ に対応した $w(t)^{(s)}$ で重みづけ平均して真の分布を近似する。

$$w(t)^{(s)} = w(t-1)^{(s)} \\ \times \frac{p(\mathbf{y}(t)|\phi(t)^{(s)}) p(\phi(t)^{(s)}|\Phi(t-1))}{q(\phi(t)^{(s)}|\Phi(t-1), \mathbf{Y}(t))}. \quad (2)$$

尤度 $p(\mathbf{y}(t)|\phi(t)^{(s)})$ の計算は [6] に準じる。

パーティクルフィルタで重要なのは proposal distribution のモデル化である。簡単には $p(\phi(t)|\Phi(t-1))$ 、すなわち単純に事前分布を用いることができる。proposal distribution を $q(\cdot)$ として、 $\Phi(t)^{(s)}$ をサンプリングする:

$$\phi(t)^{(s)} \sim q(\phi(t)|\Phi(t-1), \mathbf{Y}(t)). \quad (3)$$

前節で述べた生成モデルに従って、条件付き分布 $p(\phi(t)|\Phi(t-1))$ を次のように分解する:

$$p(\phi(t)|\Phi(t-1)) \triangleq p(c(t)|C(t-1), P_b, P_d) \quad (4)$$

$$\times \prod_m p(j_m(t)|J(t-1)) \quad (5)$$

$$\times \prod_i p(z_i(t)|Z(t-1), \gamma, \pi) \quad (6)$$

$$\times \prod_k p(\xi_k(t)|\theta_k^\xi(t-1)) p(\psi_k(t)|\theta_k^\psi(t-1)) \quad (7)$$

3.3 “Birth and Death” 変数 c

式 (4) の右辺は追跡対象の生成 (birth, addition) および消滅 (death, deletion) を表現する変数 $c(t)$ のサンプリング分布である。 $c(t)$ の時間発展は 2 段階のベルヌイ試行としてモデル化する [6]。すなわち i) 前時刻においてシーン中に存在する追跡対象 ($c_i(t-1) = 1$) は確率 P_d でシーンから消える ($c_i(t) = 0$)。そうでなければ $c_i(t) = 1$ とする。 ii) 新たな追跡対象が確率 P_b で生成される。この場合、新しいインデック \hat{i} を導入し、 $c_{\hat{i}}(t) = 1$ とする。

各時刻で新たに生成される追跡対象の数は高々 1 に制限 [6] すると、次の式を得る:

$$p(c(t)|C(t-1), P_b, P_d) \\ = P_d^{n_d} (1-P_d)^{n_s} P_b^{n_b} (1-P_b)^{1-n_b}. \quad (8)$$

ここで、 n_s はシーン中に現存している追跡対象の数、 n_d は時刻 t にシーンから消えた対象数、 $n_b \in \{0, 1\}$ が新たに生成した追跡対象の数を表す。

3.4 Data Association 変数 j

次に data association 変数 j の分布を考える (式 (5))。本モデルでは $p(j_m(t)|J(t-1))$ を一様分布、すなわち一般に data association に関して事前知識がない状況を想定する。ただし後の実験では事前知識を導入した例を示す。

J の探索空間が広い場合には、疑似的に尤度を導入することで外れ値を除外することも可能である (c.f. [13])。実験では、 j の proposal distribution として次のような分布を用いた [6]:

$$q(j_m(t) = i|J(t-1), \mathbf{y}_m(t)) \\ \propto p(\mathbf{y}_m(t)|\hat{\mathbf{x}}_i(t)) p(j_m(t) = i|J(t-1)). \quad (9)$$

上式では“代表的な”状態ベクトル $\hat{\mathbf{x}}_i(t)$ における尤度を使って事後分布に近い分布を計算する。本モデルでは $\hat{\mathbf{x}}_i(t)$ として、 $\mathbf{x}_i(t)$ 予測分布 (フィルタリング前) の平均値を利用した。

3.5 クラスタインデックス z

式 (6) における $z(t)$ の生成には Chinese Restaurant Process (CRP) [14] を利用する. CRP とは, ノンパラメトリックベイズモデルの一種である Dirichlet Process Mixture (DPM) の実現例で, サンプルの分割 (クラスタリング) の事前分布である. 理論的には可算無限個のクラスタ混合をモデル化可能であるが, 実際にはデータの複雑度に応じた数のクラスタを生成する. 本モデルでは, $Z(t-1)$ を事前の情報として, 時刻 t における $z_i(t)$ のサンプリングを次のように行う:

$$p(z_i(t) = k | Z(t-1), \gamma) = \begin{cases} \frac{m_k(t-1)}{|Z(t-1)| + \gamma} & \text{if } m_k(t-1) > 0 \\ \frac{\gamma}{|Z(t-1)| + \gamma} & \text{if } m_k(t-1) = 0. \end{cases} \quad (10)$$

$m_k(t-1)$ は時刻 $t-1$ までに得られた k 番目のクラスタの大きさである. また, $|Z(t-1)|$ は時刻 $t-1$ におけるクラスタのサイズ総数, すなわち $m_k(t-1)$ の総和である.

本モデルではダイナミクスの時間的な切替え頻度もモデル化する. 新しいパラメータ π を導入し, 確率 π でダイナミクスの変更を行うこととする:

$$z_i(t) \sim \text{CRP}(Z(t-1), \gamma) \quad \text{確率 } \pi \quad (11)$$

$$z_i(t) = z_i(t-1) \quad \text{確率 } 1-\pi \quad (12)$$

各時刻 t , 各追跡対象 i ごとに上記の確率で $z_i(t)$ を決定して適切なダイナミクスのクラスタにアサインする. したがってある対象が時刻 $t+1$ においてダイナミクスを変更した場合, $z_i(t+1) \neq z_i(t)$ となって対象の所属クラスタを変更することが可能となる.

3.6 Kalman Filter パラメータ ξ, ψ

追跡対象 i の隠れ状態 $\mathbf{x}_i(t)$ および観測量 $\mathbf{y}_m(t)$ が k 番目のダイナミクスから生成される (すなわち $j_m(t) = i$ かつ $z_i(t) = k$ のとき) 過程を Kalman filter を用いてモデル化する.

$$\mathbf{x}_i(t) = f(\mathbf{x}_i(t-1), \xi_k(t)), \xi_k(t) = \{\mathbf{q}, \mathbf{Q}\} \quad (13)$$

$$\mathbf{y}_m(t) = h(\mathbf{x}_i(t), \psi_k(t)), \psi_k(t) = \{\mathbf{r}, \mathbf{R}\}. \quad (14)$$

ここで, f と h は正規分布ノイズを持つ線形モデルである. システムノイズの平均と共分散行列を $\xi_k(t) = \{\mathbf{q}, \mathbf{Q}\}$, 観測ノイズの平均と共分散行列を $\psi_k(t) = \{\mathbf{r}, \mathbf{R}\}$ とする. これらのパラメータ推定のため, パラメータの事前分布に Normal Inverse Wishart distribution (NIW) (c.f. [8]) を仮定する. 各クラスタのシステムノイズのハイパーパラメータを θ_k^ξ , 観測ノイズのハイパーパラメータを θ_k^ψ とする. また $\theta_k = \{\theta_k^\xi, \theta_k^\psi\}$ とする.

時刻 $t-1$ における θ の事後推定値 $\theta_k(t-1) = \{\theta_k^\xi(t-1), \theta_k^\psi(t-1)\}$ を用いて, 時刻 t における真の ξ_k, ψ_k の推定値 $\xi_k(t)$ と $\psi_k(t)$ をサンプリングする. 式 (7) の各要素を次のようにモデル化してサンプリングを行う:

$$p(\xi_k(t) | \theta_k^\xi(t-1)) = \text{NIW}(\xi_k; \theta_k^\xi(t-1)) \quad (15)$$

$$p(\psi_k(t) | \theta_k^\psi(t-1)) = \text{NIW}(\psi_k; \theta_k^\psi(t-1)). \quad (16)$$

表 1 比較手法とそれぞれの特徴

method	Eqs. (17), (18)	Eqs. (10), (11), (12)
Single	no	no
Individual	yes	no
Clustered	yes	yes

サンプルされた $\xi_k(t), \psi_k(t)$ を用いて対象のトラッキングを実行した後に, トラッキングの結果からハイパーパラメータ θ のオンライン更新を行う. $\xi_k(t)$ の事後分布 ($\theta_k^\xi(t)$) は共役性から簡単に計算可能である.

$$\text{NIW}(\xi_k; \theta_k^\xi(t)) \propto p(\mathbf{x}(t) | \xi_k) \text{NIW}(\xi_k; \theta_k^\xi(t-1)). \quad (17)$$

同様に $\theta_k^\psi(t)$ に関しても次式で推定できる.

$$\text{NIW}(\psi_k; \theta_k^\psi(t)) \propto p(\mathbf{y}(t) | \psi_k, \mathbf{x}(t)) \text{NIW}(\psi_k; \theta_k^\psi(t-1)). \quad (18)$$

これらの更新済みハイパーパラメータ $\theta(t)$ は次の時刻において式 (15), 式 (16) でのサンプリングに用いられる.

4. 実験

提案法の性能を人工データと実動画データを用いた実験で確認する.

4.1 比較手法

提案モデルの比較対象として隠れ変数の数を制限した 2 モデルを用意し, ダイナミクスモデルのオンライン推定の効果を評価する. 3 モデルの特徴を表 1 にまとめる. 最初のモデル (**Single**) は固定パラメータのもとでの複数対象トラッキングモデル [6] とほぼ同様のモデルとなる. このモデルはダイナミクスのハイパーパラメータ $\theta(t)$ の更新 (式 (17), 式 (18)) を行わず, 常にデフォルトの初期値 $\theta(0)$ を用いる. さらに, CRP を用いたダイナミクスのクラスタリング (式 (10), 式 (11), 式 (12)) も行わない. 結果として全ての追跡対象に対して $\xi(t)$ と $\psi(t)$ は常に唯一つの初期分布よりサンプリングされる.

第二のモデル (**Individual**) は先節で示された $\theta(t)$ のオンライン推定を行う (式 (17), 式 (18)) が, $z_i(t)$ に対するクラスタリングを行わない. 従って, 追跡対象ごとに独立にハイパーパラメータを学習するが, その数は 1 対象につき 1 つだけである. このモデルでは, 最初のモデルにくらべ複数の追跡対象が異なるダイナミクスを持つことを表現できる点で複雑になっているが, パラメータのパターン化 (クラスタリング) は行われぬ. これは, 1 対象から得られるダイナミクスの情報は少ない (高々フレーム長) ため, さらにクラスタに分割すると学習に必要な情報が少なくなると思われるためである.

最後に提案モデル (**Clustered**) では, 先章までに説明した通り, ダイナミクスのクラスタリングとそれぞれのハイパーパラメータをオンラインで推定する. 第二のモデルとの相違点は, 各追跡対象が共有した運動パターンを用いて軌跡を生成するようにモデル化している点にある. 本モデルでは, 対象固有のダイナミクスを推定する代わりに, 複数対象間で推定に用いる情報を集めることでクラスタ化されるダイナミクスの学習を安定

表 2 人工データ生成に用いられたパラメータのパターン (4 種類). 各パターンにおいて $\mathbf{r} = \mathbf{0}$ である

	\mathbf{q}	\mathbf{Q}	\mathbf{R}
I	$\{3.0, 0.0\}^T$	$\text{diag}\{1.0, 1.0\}$	$\text{diag}\{1.0, 3.0\}$
II	$\{-3.0, 0.0\}^T$	$\text{diag}\{1.5, 1.5\}$	$\text{diag}\{1.0, 2.0\}$
III	$\{0.0, 3.0\}^T$	$\text{diag}\{0.5, 0.5\}$	$\text{diag}\{2.0, 2.0\}$
IV	$\{0.0, -10.0\}^T$	$\text{diag}\{1.0, 1.0\}$	$\text{diag}\{0.5, 3.5\}$

させる.

4.2 人工データを用いた実験

この実験では, $[0 : 200] \times [0 : 200]$ の仮想 2 次元空間中の移動質点のトラッキングとクラスタリングをタスクとする. 追跡対象 (質点) の隠れ状態は, 各対象の位置を表す 2 次元ベクトルとする. また, 観測量も 2 次元ベクトルである. これらはノイズによって劣化した対象位置の情報を表す. また, 観測データにはどの追跡対象にも属さないランダム観測も含まれる. 各対象は次のランダムウォークモデルに従う:

$$\mathbf{x}_i(t) = \mathbf{x}_i(t-1) + \mathbf{v}(t), \mathbf{v}(t) \sim N(\mathbf{q}, \mathbf{Q}) \quad (19)$$

$$\mathbf{y}_i(t) = \mathbf{x}_i(t) + \mathbf{w}(t), \mathbf{w}(t) \sim N(\mathbf{r}, \mathbf{R}). \quad (20)$$

ここで $N(\cdot)$ は正規分布を表す.

各対象は表 2 に示す上下左右への移動に対応する 4 つのダイナミクスパターンをランダムに切替えながら軌跡と出力を生成するものとした. なおすべてのパターンで $\mathbf{r} = \mathbf{0}$ である. NIW の初期値 $\theta(0)$ は, $\xi_k(t)$ および $\psi_k(t)$ の平均値が $\mathbf{q} = \mathbf{0}$, $\mathbf{Q} = \text{diag}\{10.0, 10.0\}$, $\mathbf{r} = \mathbf{0}$, $\mathbf{R} = \text{diag}\{5.0, 5.0\}$ となるように設定する. 時系列データは 300 ステップである. パーティクル数は $S = 300$, CRP の concentration parameter は $\gamma = 2$ とした. 同時に追跡される対象数は 10 対象程度である. 各時刻において $P_b = 0.1$ の確率で新しい追跡対象が発生するものとする. シーン中に存在する対象 ($c_i(t-1) = 1$) は式 (21) で計算される確率でシーンから消滅するものとする (追跡対象の消滅確率 P_d):

$$P_d = 1 - \lambda e^{-t/t_n}. \quad (21)$$

t_n は $j_m(\cdot) = i$ となる観測量 \mathbf{y}_m が少なくとも一つ存在した最後の時刻を表す ($\lambda = 0.1$ とする).

続いて実験結果を検討する. データ対数尤度の時刻平均値を表 3 の第 3 列に示す. 表より, Kalman filter のパラメータ推定およびそのクラスタリングが対数尤度を向上することが確認できる. これは, 提案モデルがオンラインのトラッキングとクラスタリングによってより良いモデルとパラメータを獲得したことを示している.

図 3 に示したのは提案法 ($\pi = 0.2$) における, 正規分布ノイズのパラメータ \mathbf{q} の最終時刻における分布である. \mathbf{q} は追跡対象の平均速度バイアスを表しており, 各プロット点は一つのダイナミクスクラスタの \mathbf{q} の平均値を表している. 図の Clst0~9 は推定されたクラスタを表す. “size” で示された数字は各クラスタのデータサイズ (どれだけ多くの対象がそのクラスタに属したか) を表す. 4 つの主要なクラスタ (size > 100) が獲得され,

表 3 人工データおよび実動画データ実験における平均対数尤度

Method	π	synthetic	real movie 1	real movie 2
Single	-	-115.218	-95.617	-54.543
Individual	-	-107.850	-84.205	-51.937
Clustered	0.1	-96.678	-79.710	-50.297
	0.2	-96.369	-80.925	-50.740
	0.5	-103.659	-82.957	-51.857
	1.0	-110.313	-83.931	-52.130

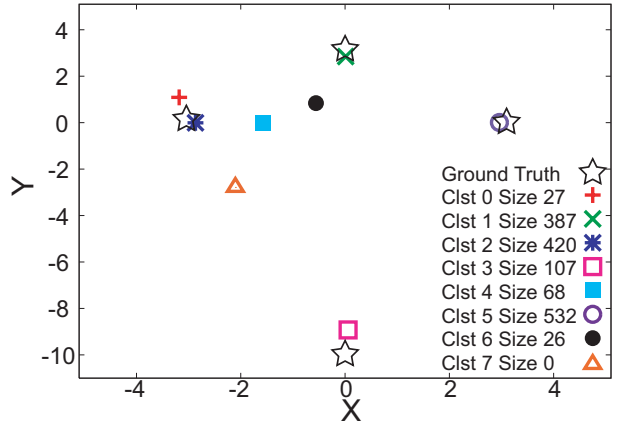


図 3 人工データ実験における速度ノイズの平均ベクトル分布. 星印で示された真値 (Ground Truth) に近い位置に, 大きなクラスタが形成されていることが分かる

それらは人工データ生成時に用いられた真値 (Ground Truth) に近い値を得た. すなわち, 提案モデルはダイナミクスパターンのクラスタリングとそれらのパラメータ推定に成功したといえる.

4.3 実動画データによる実験



図 4 実動画データの実験で用いられたデータの例. (a) は第一の実験, (b) は第二の実験で用いた動画である.

次に, 2 種類の実動画データを用いた実験について説明する. デジタルカメラで撮影された動画を用いて, シーン中の歩行者から抽出される特徴点の追跡とクラスタリングを行う.

第一の実験では, ほぼ直上から撮影された歩行者から抽出した特徴点を追跡する (図 4(a)). 各フレームの大きさは 320×240 ピクセルである. 動画データを観察すると, まれに店に入るため左右への方向転換が観察されたが, 基本的な動きのパターンは画面上下方向の 2 種であった.

目標となる特徴点は次のようにして抽出した. 最初に背景差分を行い, 二値化によって前景ピクセルを抽出する. 続いて, 前

景部分中から黒色のピクセルだけを選び、それらを mean shift クラスタリングで少数のクラスタ中心へ量子化する。これらのクラスタ中心座標を特徴点とする。おおよそ一人の歩行者から 1～3 個の特徴点が抽出された。

本実験では j の分布 (式 (5)) に色モデルを設定した [4]。各特徴点の周辺ピクセルにおける RGB ヒストグラムと、前時刻に得られた追跡対象の RGB ヒストグラムとの Bhattacharya 係数から association の確率を計算する。対象 i のヒストグラムは、 $j_m(t-1) = i$ となった観測点のヒストグラムの平均とする。 $j_m(t) = i$ となる確率を次式で定義した:

$$P(j_m(t) = i | J(t-1)) \propto \exp\left(-\frac{\|d(m, i)\|^2}{\sigma^2}\right).$$

ここで σ は事前に設定した定数パラメータ ($\sigma = 0.15$) で、 $d(m, i)$ は m 番目の観測値と i 番目の追跡対象のヒストグラム間の Bhattacharya 係数である。

状態空間モデルは先の人工データ実験と同じものを用いる。NIW の初期値 θ_0 は、 $\xi_k(t)$ および $\psi_k(t)$ の平均値が $\mathbf{q} = \mathbf{0}$ 、 $\mathbf{Q} = \text{diag}\{20.0, 20.0\}$ 、 $\mathbf{r} = \mathbf{0}$ 、 $\mathbf{R} = \text{diag}\{20.0, 20.0\}$ となるように設定した。動画のフレーム数は 200 フレーム (6FPS) である。パーティクルの総数は $S = 500$ 個、CRP パラメータは $\gamma = 0.1$ とした。新規追跡対象の生成確率は $P_b = 0.1$ 、追跡対象の消滅確率は式 (21) で計算する ($\lambda = 0.1$)。

第二の実験では、混雑した状況下での歩行者追跡をタスクとした。先般の実験にくらべてより低い視点から交差点での歩行者を撮影したため、多くのオクルージョンが発生するデータになっている (図 4(b))。各フレームの大きさは 640×480 ピクセルである。動画データを観察すると、ほとんどの歩行者は画面内の上下方向、あるいは右から左へと歩いているため、獲得される運動クラスタも上、下、および左方向へのバイアスがかかったモデルが予想される。ただし、混雑での衝突を回避するため、細かい運動方向の変化も観察された。

目標となる特徴点は、Histogram of Gradient 特徴を用いた人物検出法 [15] を利用して抽出する。著者らが配布しているプログラムを撮影した動画に適用し、各フレームから検出された人物形状の位置座標を特徴ベクトルとして利用した。オクルージョンが起ると検出に失敗しやすいためにミッシングデータが多数存在するデータとなった。

状態空間モデルはこれまでの実験と同じものを用いる。NIW の初期値 θ_0 は、 $\xi_k(t)$ および $\psi_k(t)$ の平均値が $\mathbf{q} = \mathbf{0}$ 、 $\mathbf{Q} = \text{diag}\{25.0, 25.0\}$ 、 $\mathbf{r} = \mathbf{0}$ 、 $\mathbf{R} = \text{diag}\{10.0, 10.0\}$ となるように設定した。動画のフレーム数は 300 フレーム (7.5FPS) である。パーティクルの総数は $S = 1000$ 個、CRP のパラメータ $\gamma = 1$ とした。新規追跡対象の生成確率は $P_b = 0.1$ とし、追跡対象の消滅確率は式 (21) で計算する ($\lambda = 0.1$)。

続いて実験結果を検討する。表 3 の右 2 列に計算された平均対数尤度を示す。人工データの実験の場合と同様、提案法によって対数尤度が向上することが確認できる。対数尤度が高いということは、モデルがデータに良く適合しているということになるので、精度の高い追跡が期待できる。次に、平均速度の

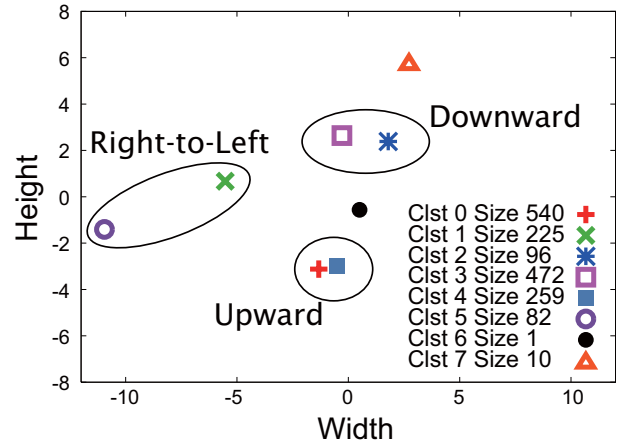


図 5 実動画データ実験における速度ノイズの平均値分布 (実験 2)。大まかに上方向、下方向、および左方向への運動モデルが獲得された。図中の楕円によるクラスタリングは著者らによる

バイアスである \mathbf{q} の分布を図 5(実験 2) に示す。実データであるため ground truth は存在しないが、上下方向、そして左向きの速度バイアスをもつ少数の dominant なダイナミクスのクラスタに分離している。またそのクラスタも上方向、下方向、左方向へ速度ベクトルのバイアスがあるモデルであり、おおよそ動画の観察から予想されたとおりのクラスタが得られた。

最後に、第実動画データ実験におけるトラッキング結果の連続スナップショットを図 6(実験 1)、図 7(実験 2) に示す。対象の推定位置は色の付けた矩形で示している。それぞれの矩形の色は追跡対象の ID(i) に対応する。また、矩形左上の数字はダイナミクスクラスタのインデックス (z) に対応する (図 5)。混雑した状況下でも、追跡対象のダイナミクスを区別した追跡がある程度まで実現できていることがわかる。

5. む す び

複数のダイナミクスモデルで駆動される複数対象のトラッキングをモデル、対象数とも未知の状態で解決する確率的生成モデルを提案した。提案法では、ダイナミクスの混合モデルの混合数と各コンポーネントのパラメータをノンパラメトリックベイズモデルの枠組で推論した。これによって、複数対象のトラッキングと、これらの対象が従うダイナミクスのパターンを同時に学習するモデルが実現された。人工データと実動画データを用いた実験によって、提案法の有効性を確認した。

今後は、線形ガウシアンモデルよりも更に複雑なモデルを組み込んで、物体のアピランスや構造のオンライン学習にも発展させたいと考える。

文 献

- [1] A. Mittal and L. S. Davis: "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo", International Journal of Computer Vision, **51**, 3, pp. 189-203 (2002).
- [2] Z. Khan, T. Balch and F. Dellaert: "A Rao-Blackwellized particle filter for eigentracking", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 980-986 (2004).
- [3] Y. Li, H. Ai, T. Yamashita, S. Lao and M. Kawade: "Track-

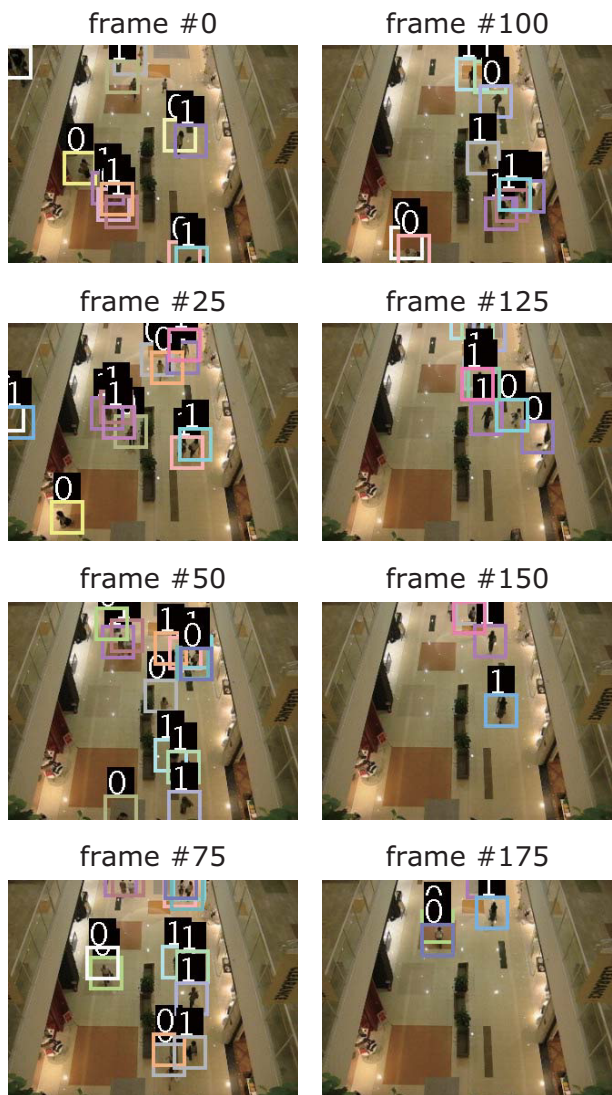


図 6 追跡結果の例 (第一の実動画画像実験). 矩形の色は追跡対象の ID(i), 上部の数字がダイナミクスのインデックス (z) に対応する. この例では, ダイナミクスはほぼ画面垂直方向の動きのみであり, 下方向 (1 番) と上方向 (2 番) のクラスタだけでほぼ全ての動きが表現された

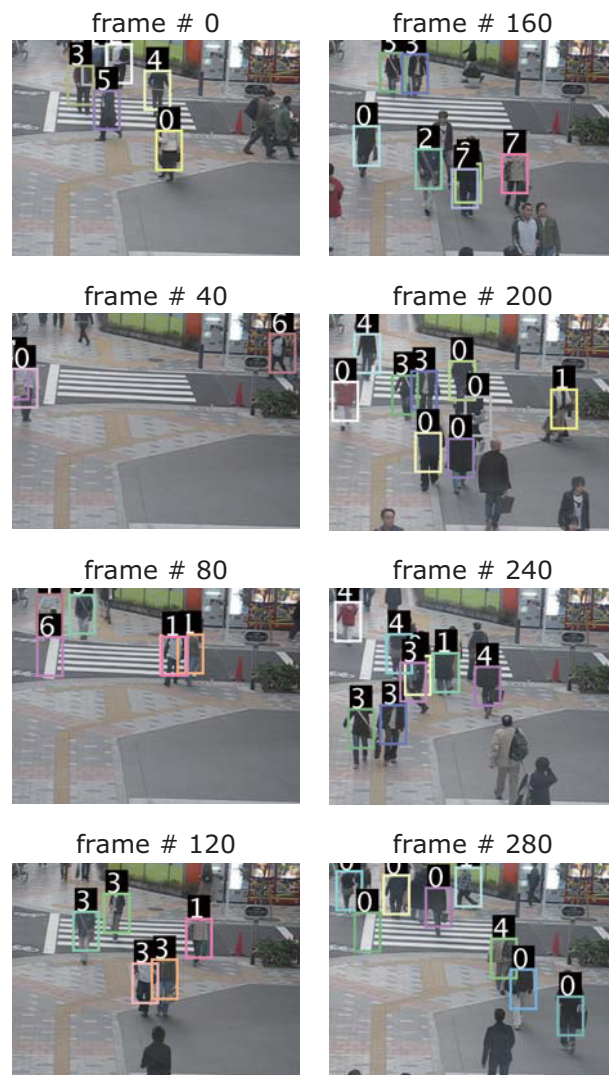


図 7 追跡結果の例 (第二の実動画画像実験). 矩形の色は追跡対象の ID(i), 上部の数字がダイナミクスのインデックス (z) に対応する. 各ダイナミクスの詳細については図 5 を参照

ing in low frame rate video: A cascade particle filter with discriminative observers of different lifespans”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, USA (2007).

- [4] B. Leibe, N. Cornelis, K. Cornelis and L. V. Gool: “Dynamic 3D scene analysis from a moving vehicle”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, USA (2007).
- [5] Z. Khan, T. Balch and F. Dellaert: “MCMC data association and sparse factorization updating for real time multi-target tracking with merged and multiple measurements”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**, 12, pp. 1960–1972 (2006).
- [6] S. Särkkä, A. Vehtari and J. Lampinen: “Rao-Blackwellized particle filter for multiple target tracking”, Information Fusion, **8**, 1, pp. 2–15 (2007).
- [7] J. Ting, E. Theodorou and S. Schaal: “A Kalman filter for robust outlier detection”, Proceedings of IEEE International Conference on Intelligent Robotics Systems, San Diego, USA, pp. 1514–1519 (2007).
- [8] F. Caron, M. Davy, A. Doucet, E. Duflos and P. Vanheeghe:

“Bayesian inference for linear dynamic models with Dirichlet process mixtures”, IEEE Transactions on Signal Processing, **56**, 1, pp. 71–84 (2008).

- [9] K. Ishiguro, T. Yamada and N. Ueda: “Simultaneous clustering and tracking unknown number of objects”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, USA, p. to appear (2008).
- [10] M. S. Arulampalam, S. Maskel, N. Gordon and T. Clapp: “A tutorial on particle filters for online nonlinear / non-gaussian bayesian tracking”, IEEE Transactions on Signal Processing, **50**, 2, pp. 174–188 (2002).
- [11] 北川: “時系列解析入門”, 岩波書店 (2005).
- [12] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada and N. Ueda: “Learning systems of concepts with an infinite relational model”, Proceedings of the 21st National Conference on Artificial Intelligence (2006).
- [13] M. K. Pitt and N. Shephard: “Filtering via simulation: Auxiliary particle filters”, Journal of the American Statistical Association, **94**, 446, pp. 590–599 (1999).
- [14] D. Blackwell and J. B. MacQueen: “Ferguson distributions via Polya urn schemes”, The Annals of Statistics, **1**, 2, pp. 353–355 (1973).
- [15] N. Dalal and B. Triggs: “Histograms of oriented gradients

for human detection”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, San Diego, USA, pp. 886–893 (2005).