深層学習を導入したデータ同化による複数物体追跡手法の構築

石井　健太[1]・瀬尾　亨[2]・布施　孝志[3]

[1]非会員　東京大学大学院　工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）
E-mail: ishii@trip.t.u-tokyo.ac.jp
[2]正会員　東京大学大学院　工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）
E-mail: seo@civil.t.u-tokyo.ac.jp
[3]正会員　東京大学大学院　工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）
E-mail: fuse@civil.t.u-tokyo.ac.jp

歩行者や車両の軌跡は空間設計や流動制御を行うための重要なデータとなる．このため，さまざまな物体追跡手法が提案されており，特に観測データとシミュレーションを統合するデータ同化による手法が盛んに研究されている．しかしながら，データ同化を用いた追跡手法の多くは，特徴量を手動で決定した観測モデルを用いているため利用可能性が限定されるという課題が残存する．一方で，画像認識分野では Convolutional Neural Network（CNN）を用いた認識器が高い精度を誇っており，これを用いた新たな追跡手法の構築が期待される．本論文は，データ同化の枠組みの中で CNN を観測モデルとして導入した複数物体追跡手法の構築を目的とする．まず，基本手法としてパーティクルフィルタを用いる．そして，上記手法に考えられる課題を解決するため，Probability Hypothesis Density（PHD）フィルタへ拡張した手法を構築する．提案した手法を実データに適用し，その性質を検証した．

Key Words: *data assimilation, convolutional neural network, particle filter, PHD filter, fine tuning*

1. はじめに

歩行者や車両の軌跡データに基づく施設設計・流動制御が注目されている．軌跡データを取得する方法は数多く存在するが，その中でもカメラは多くの場所に設置されるようになってきており，視野内であれば全数調査が可能であるという利点も持つ．しかしながら，画像による物体追跡はオクルージョンや影による見えの変化に影響を受けるといった問題も存在する．この問題を解決するために，見えから得られる情報だけでなく，物体の挙動に関するモデルを用いた予測を組み合わせるデータ同化を適用した手法が存在する[1)]．この手法により，見えの変化など観測のみでは追跡が難しかった状況下においても，システムモデルによって前時刻の状態ベクトルとの関連性を考慮できるようになり，追跡精度が向上している．しかしながら，データ同化を用いた追跡手法の多くは，手動で決定した観測モデル（例：カラーヒストグラムの相関係数である Bhattacharyya 係数を用いる手法）を用いているため，利用可能性が限定されているという課題が残存する．

画像処理分野における研究に着目すると，機械学習の一種である Convolutional Neural Network (CNN) が高い精度を誇っている[2)]．この理由としては，主に以下の 2 つが挙げられる．1 つ目は，抽出する特徴量をヒューリスティックに決定するのではなくデータに基づく学習により決定することである．2 つ目には，ネットワークに深い層を導入することにより，非線形性が強い問題に対しても自由度の高いモデリングが可能であることが挙げられる．しかしながら CNN は認識器であるため，物体追跡に適用するためにはほかの手法を組み合わせる必要がある．このため，CNN のもつ画像認識精度と，物体追跡手法の枠組みとしてのデータ同化を組み合わせた新たな物体追跡手法が期待される．

本研究の目的は，データ同化の枠組みの中で CNN を観測モデルとして導入した複数物体追跡手法を構築することである．本稿の構成は以下の通り．**第 2 章**にてデータ同化を用いた物体追跡手法を概観し，従来手法の長所と課題を整理する．その後，CNN を用いた画像認識手法，物体追跡手法についても同様に整理を行う．**第 3 章**にて CNN をデータ同化の計算手法の 1 つであるパーティクルフィルタに導入する枠組みを提示する．**第 4 章**にて**第 3 章**で提示した手法に予測される課題を解決するため，Probability Hypothesis Density(PHD) フィルタへ拡張した手法を提示する．**第 5 章**にて**第 3 章**，**第 4 章**で提示した手法を実データに適用し，それらの性質を検証する．**第 6 章**にて結論をまとめる．

2. 既往研究の整理

(1) データ同化を用いた物体追跡手法

データ同化を用いた物体追跡手法を歩行者に適用した例としては，布施・中西[1]の研究が挙げられる．この研究では動きの情報として離散選択モデルによる歩行者挙動モデルを，見えの情報として色情報を用いたBhattacharyya係数と距離情報を併用している．中西[3]ではPlan-Viewと呼ばれる床平面を格子状に分割した2次元マップを用いて人物候補の抽出手法と初期分布の設定方法を検討するほか，人物追跡を行うためのシステムをランダムウォーク，等速直線運動，離散選択モデルの3つから最適なものを逐次的に選択する手法を提案している．一方で，車両に適用した例として，視野を共有する複数カメラを用いた追跡手法も存在する．[4]この手法では物体追跡で問題となる空間的な誤差に加えて，カメラ間の時刻が非同期であることによる時間的誤差を考慮しているほか，見えの情報から複数カメラ間で各車両を結び付けている．

上記手法はいずれも逐次ベイズフィルタの枠組みを用いており，実装にはパーティクルフィルタを利用している．これらは状態ベクトルを物体の位置とし，単体の物体を追跡する枠組みとして用いられることが多い．このアプローチを複数物体の追跡に適用する際には，状態ベクトルの大きさが人数に比例するため，状態空間全体をパーティクルで網羅するためには計算量が指数関数的に増大してしまうといった問題点が存在する．また，対象を検出できないケースや偽の検出をしてしまうケースの考慮が煩雑になる．この問題を解決する複数物体追跡のための新たな手法としてランダム有限集合の枠組みを用いたPHDフィルタが考案された[5]．本手法では，人数とその軌跡を同時に最適化する．PHD(確率仮説密度)とは推定すべき事後分布を1次モーメントを用いて近似した分布であり，ある領域でPHDを積分するとその領域内に存在する対象物体数の期待値となる．1次のモーメントのみを推定することで，上記で述べた対象の個数が確率的に変化する状態空間モデルにおいて状態推定を行った場合，推定する確率分布の標本空間が複雑になってしまう問題を解決する手法となっている．

しかしながら，PHDフィルタも逐次ベイズ推定同様，実装のためには計算を可能にするための仮定を置く必要がある．Voら[6]はPHDフィルタを逐次モンテカルロ法の枠組みで実装する手法(SMC-PHDフィルタ)を提案している．またRisticra[7]はそのアルゴリズムを改善し，新規出現と残存を区別可能にしている．一方で，線形性と正規性を仮定することによりPHDフィルタにおけるフィルタリング式をclosed-formにする手法[8]も提案されている．また生駒ら[9]は，このSMC-PHDフィルタにニューラルネット(NN)を観測モデルに用いた手法とNNと空間相関のみで追跡する手法の精度比較を行っている．

(2) 深層学習を用いた物体認識・追跡手法

画像認識分野においては，前述したようにCNNが高い精度を誇っている．このため，まずCNNを用いた一般物体認識モデルについて概観する．一般物体認識モデルの精度向上はILSVRC[2]と呼ばれる大規模画像認識コンペティションの最優秀モデルを辿るとわかりやすい．まず，2012年にAlexNet[10]と呼ばれるCNNモデルが優勝し，初めてCNNを用いたモデルが優勝した年となった．これは畳み込み層5層，全結合層3層からなるモデルとなっている．活性化関数としてReLUが用いられており，また学習時にDropoutを導入しているなど，最新のモデルにも標準的に採用されている手法を先駆けて利用しているモデルである．以降CNNを用いたモデルが優勝するようになり，2014年には，GoogLeNet[11]と呼ばれるモデルが登場した．このモデルの特徴は，Inceptionモジュールと呼ばれる小さなネットワークを定義し，これ自体を畳み込み層のように重ねている点である．このInceptionモジュールにより，畳み込み層の重みをスパースにし，パラメータ数とのトレードオフを改善している．このGoogLeNetは以降も畳み込み層のサイズを変化させたバージョンがいくつか出てきている．

次に深層学習を用いた追跡手法としては，上記で述べたCNNを用いたものや，時系列特徴量抽出に適したRecurrent Neural Network（RNN）を用いたものが存在する．前者の例としては，CNNの，出力側に近いもので，位置に関する情報はぼやけているが分類には適した特徴量と，入力側に近いもので，位置に関する情報を色濃く持っているが分類可能なまでに特徴を抽出しきれていない特徴量を組み合わせることで，画面内の物体の追跡を行う手法が存在する[12]．後者の例としては，各時刻の各物体を相関づける相関フィルタに着目し，これをオンライン学習により逐次更新していくために，RNNの枠組みを用いて広範囲を参照し信頼性のあるバッチを抽出する手法が存在する[13]．しかしながら複数物体追跡においては，CNN単体を用いた手法では難しく，またRNNなどのオンライン学習では計算コストが大きくなってしまう．深層学習の特徴量抽出に関する性能が高いという特長は生かしながらも，ほかの手法と組み合わせた追跡手法が求められる．

3. 基礎手法の構築

まず基礎手法として，一般状態空間モデルの枠組みの中で，観測モデルに CNN を導入した複数物体追跡手法を構築する．実装手法としてはパーティクルフィルタ（PF: Particle Filter）[14] を用いる．

(1) 一般状態空間モデル

一般状態空間モデルは

$$\boldsymbol{x}_t \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \tag{1}$$

$$\boldsymbol{y}_t \sim p(\boldsymbol{y}_t|\boldsymbol{x}_t) \tag{2}$$

と表せる．ここに，$\boldsymbol{x}_t$ は時刻 t における対象物体の潜在状態を意味する状態ベクトル，$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ は対象物体の挙動が従うシステムモデル，$\boldsymbol{y}_t$ は時刻 t における観測値を意味する観測ベクトルである．$p(\boldsymbol{y}_t|\boldsymbol{x}_t)$ は観測値の得られ方を意味する観測モデルである．

実装手法として PF を用いることで，上記のように線形性やガウス性の仮定を置かない一般状態空間モデルにおいても，$\boldsymbol{x}_t$ の事後分布 $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ を逐次的に計算可能である．解法は参考文献[14] を参照されたい．

(2) システムモデル

本手法では，実データとして車両・歩行者両方に適用することを見据え，システムモデルに等速直線運動を用いる．このため，状態ベクトルを

$$\boldsymbol{x}_t = \begin{pmatrix} x_t \\ y_t \\ \boldsymbol{v}_t \end{pmatrix} \tag{3}$$

とする．ただし，x_t, y_t は時刻 t の状態空間における x, y 座標であり，$\boldsymbol{v}_t = (v_{x,t}, v_{y,t})^{\mathrm{T}}$ は時刻 t における追跡対象の持つ速度ベクトルである．以上よりシステムモデルは式 (4) のように表せる．

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = p\begin{pmatrix} x_{t-1} + v_{x,t}|x_{t-1}, y_{t-1}, \boldsymbol{v}_{t-1} \\ y_{t-1} + v_{y,t}|x_{t-1}, y_{t-1}, \boldsymbol{v}_{t-1} \end{pmatrix} \tag{4}$$

また速度 $\boldsymbol{v}_t$ については上述した位置に関するシステムモデルの入力となっている．速度に関するシステムモデルは，前時刻の速度を平均とした標準偏差 $\boldsymbol{\sigma}$ の正規分布に従うこととする．

(3) CNN を用いた観測モデルの作成

観測ベクトルを

$$\boldsymbol{y}_{i,j,t} = \begin{pmatrix} \boldsymbol{X}_{i,j,t} \\ R_{i,j,t} \\ G_{i,j,t} \\ B_{i,j,t} \end{pmatrix} \tag{5}$$

とする．ただし，i, j は画面内の画素の index であり，$\boldsymbol{X}_{i,j,t}$ は撮影領域における座標，R, G, B はそれぞれ 0-255 の 256 段階で表される赤，緑，青色の強さを表している．

ここで，観測モデルに CNN を用いる手法について説明する．観測モデルは，対象の状態ベクトルの予測値から観測ベクトルを参照し，CNN に入力した際の出力を尤度として用いるものとする．具体的には，あらかじめウィンドウサイズを設定しておき，状態ベクトルが示唆する観測空間における部分画像を CNN の入力とする．

本手法では既存の CNN モデルを転移学習することで人物および車両の認識モデルをそれぞれ作成した．既存モデルには，**第 2 章**で触れた GoogLeNet のフィルタのサイズを変更した Inception v3 というモデルを採用した．理由としては，Inception モジュールにより認識精度とパラメータ数のトレードオフを改善したモデルとなっていることが挙げられる．転移学習により全結合層のパラメータのみを学習させた．データセットとしては，人物用のモデルには人物クラスと人が写っていない背景クラスに分かれているデータセット INIRIA Person Dataset[15] を用いた．車両用のモデルには Vehicle Make and Model Recognition Datast(VMMRdb)[16] という 1950 年から 2016 年までに製造・発売された自動車の画像が，メーカー・モデル・年代の 3 階層で分類されているデータセットを利用した．車のクラスとしては VMMRdb の車両の画像を，背景のクラスとして上述した INIRIA Person Dataset の背景クラスを用いてデータセットを新たに作成した．人物認識モデル作成時は人物クラス 900 枚，非人物クラス 1800 枚を，訓練データとテストデータの割合を 0.8 として学習に用いた．車両認識モデル作成時には各クラス 5000 枚ずつを，訓練データとテストデータの割合を 0.8 として学習に用いた．各モデルのハイパーパラメータを**表- 1** に，上記の設定の下で学習させたモデルの性能を**表- 2** に示す．ここに，Loss の定義はパラメータ学習時に使用した損失

表–1 各モデルのハイパーパラメータ

	車両	人物
Epoch	20	30
Batch size	32	64
Learning rate	1×10^{-4}	1×10^{-3}

表–2 学習後テストデータ適用時の Loss と Accuracy

	Loss	Accuracy
人物認識モデル	0.1187	0.9722
車両認識モデル	0.008127	0.9966

関数の出力であり，Accuracy の定義は正解率 (=正解数/テスト回数) である．

4. 拡張手法の構築

基礎手法は，車両など動きをシステムモデルで説明しやすい対象を追跡する場合には高い追跡精度を見込むことができる．しかし，システムモデルにおいて確率が低い移動が発生した際，確率が高い移動先に同種の物体が存在した場合，CNN が「追跡中の対象らしさ」ではなく「人らしさ」を基準に尤度の判定を行うことが理由で，追跡誤差が大きくなってしまう課題が予想される．また，第 2 章でも述べたが，パーティクルフィルタは状態空間において逐次的な次元の増減を許容しないため，追跡対象数が増加すると計算量が指数関数的に増加してしまうという課題も存在する．

これらの課題に対処するため，基礎手法を PHD フィルタを用いた複数物体追跡手法へ拡張する．これは，対象物体の位置と共に，その数も推定可能な分布である PHD を逐次推定する手法であり，人数とその位置の同時推定が可能である．

(1) ランダム有限集合に基づく状態空間モデル

ランダム有限集合に基づく状態空間モデルは

$$\boldsymbol{\Xi}_t = \boldsymbol{S}(\boldsymbol{X}_{t-1}) \cup \boldsymbol{\Gamma}(\boldsymbol{X}_{t-1}) \tag{6}$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{E}(\boldsymbol{X}_t) \cup \boldsymbol{C}_t \tag{7}$$

と表せる．ここで，$\boldsymbol{\Xi}_t$ は時刻 t における対象数を可変とした際の状態であり，複数対象の状態ベクトルを表した $\boldsymbol{X}_t$ のランダム有限集合となっている．$\boldsymbol{S}(\boldsymbol{X}_{t-1})$ は，前時刻からの残存を表し，$\boldsymbol{\Gamma}(\boldsymbol{X}_{t-1})$ は新規出現を表したシステムモデルとなっている．また $\boldsymbol{\Sigma}_t$ は時刻 t における対象数を可変とした際の観測であり，複数対象の観測ベクトルを表した $\boldsymbol{Y}_t$ のランダム有限集合となっている．$\boldsymbol{E}(\boldsymbol{X}_t)$ は，追跡対象の観測を表す項であり，$\boldsymbol{C}_t$ は誤検出を表す．実装手法として PHD フィルタの逐次モンテカルロによる近似的実装を用いることで，上記のような複数対象を扱う状態空間モデルにおいても，線形性やガウス性を仮定せず，PHD の事後分布を逐次推定することができる．解法は参考文献[6] を参照されたい．本手法では，上記手法における各粒子に対して，個人を明示的に識別するためのラベルを付与する手法[17] を用いる．ラベルの付与に関しては，実装手順とともに後述する．

(2) 状態と観測の定義

基礎手法と同様に各対象の状態ベクトルを式 (8) のように定義し，状態を式 (9) と定義する．

$$\boldsymbol{x}_{n(k),t} = \begin{pmatrix} x_{k,n(k)} \\ y_{k,n(k)} \\ \boldsymbol{v}_{k,n(k)} \end{pmatrix} \tag{8}$$

$$\boldsymbol{X}_k = \{\boldsymbol{x}_{k,1}, \boldsymbol{x}_{k,2}, \cdots, \boldsymbol{x}_{k,n(k)}\} \subset E_s \tag{9}$$

観測は

$$\boldsymbol{Y}_{k,i} = \{ROI(\boldsymbol{I}_k, \boldsymbol{x}_{k,1}), \cdots, ROI(\boldsymbol{I}_k, \boldsymbol{x}_{k,m(k)}), \boldsymbol{I}_k\} \subset E_o \tag{10}$$

とする．ここで，$ROI(\boldsymbol{I}_k, \boldsymbol{x}_{k,i})$ は観測フレーム $\boldsymbol{I}_k$ において各状態ベクトル i が示唆する部分領域内の画像を意味する．起こりうる観測状態を記述した $\boldsymbol{\Sigma}_k = \{\boldsymbol{Y}_{k,1}, \boldsymbol{Y}_{k,2}, \cdots, \boldsymbol{Y}_{k,M(k)}\}$ から確率 p_D で $\boldsymbol{Y}_{k,i}$ を取得するものとする．

個々の対象に関するシステムモデルと観測モデルは基礎手法と同様のものを用いる．

(3) 逐次モンテカルロフィルタ

a) 初期分布生成

1 人当たりの粒子数を ρ とする．まず状態空間上に一様分布に従い粒子を生成する．この粒子を観測モデルを用いて重みづけし，リサンプリングを行う．これを初期分布とする．ラベルに関しては，個数のみ所与とし，k-means 法によりクラスタリングした結果を仮のラベルとして与える．

b) 残存粒子の次時刻への伝搬

前時刻から残存する粒子をシステムモデルを用いて伝播する．このとき残存率を p_s とし，重みを $\tilde{\boldsymbol{w}}_k^{(i)} = p_s \boldsymbol{w}_{k-1}^{(i)}$ として更新する．

c) 新規出現粒子

PHD フィルタはポアソン過程に従い規定されているため，新規出現対象の個数 $N_\Gamma(k)$ は平均 μ_Γ のポアソン分布に従うとする．また新規出現位置に関しては，下記のようなグリッド状に粒子 140 個を捲くことで新規出現に関するプロポーザル分布を担わす．このため，各粒子の重みを $N_\Gamma(k)/140$ とする．

d) 重みの更新

逐次モンテカルロの近似的実装では，フィルタリングによる重みの更新を式 (11) で表す．

$$\hat{w}_k^{(i)} = \{(1 - p_D) + p_D \Psi(\boldsymbol{Y}_k, \boldsymbol{x}_k)\} \tilde{w}_k^{(i)} \tag{11}$$

ただし尤度の項 $\Psi(\boldsymbol{Y}_k, \boldsymbol{x}_k)$ は

$$\Psi(\boldsymbol{Y}_k, \boldsymbol{x}_k) = \sum_{\boldsymbol{y} \in \boldsymbol{Y_k}} \frac{h(\boldsymbol{y}|\boldsymbol{x}_k)}{\mu_C p_C(\boldsymbol{y}) + C_k(\boldsymbol{y})} \tag{12}$$

と表す．ここで畳み込みの部分にあたる $C_k(\boldsymbol{y})$ は式 (13) のように重み付き平均尤度となる．

$$C_k(\boldsymbol{y}) = \sum_{i=1}^{L_{k-1}+J_k} \tilde{w}_k^{(i)} h(\boldsymbol{y}|\tilde{\boldsymbol{x}}_k^{(i)}) \tag{13}$$

まず粒子自体の尤度は $h(ROI(\boldsymbol{I}_k, \tilde{x}_k^{(i)})|\tilde{x}_k^{(i)})$ と表すことができる．ここで注意しなければならないのは，粒子の重みと観測モデルから算出される尤度が異なるという点である．次にフィルタリング計算について詳しく述べる．まずラベルごとに $C_k(\boldsymbol{y})$ を計算する．この時，観測モデルは粒子が示唆する部分画像を参照し，その画像を CNN モデルを入力，その尤度を出力とするモデルである．このため $C_k(\boldsymbol{y})$ は 1 期先予測を通して伝わった重み $\tilde{\boldsymbol{w}}_k^{(i)}$ と $h(ROI(\boldsymbol{I}_k, \tilde{x}_k^{(i)})|\tilde{x}_k^{(i)})$ の積をとり，それらを対象ごとに和をとったものとなる．また本手法では，偽の検出を考慮する項である式 (12) の分母第 1 項については，$p_C(\boldsymbol{y})$ は観測空間の全面に一様に偽の検出が起こりうる可能性があると考え，一様分布を採用した．

e)　ラベルの付与・削除

まず，各粒子の尤度に閾値 $\beta_{likelihood}$ を設け，これを下回った粒子のラベルを 0 とする．次に粒子の位置に閾値 $\beta_{position}$ を設け，各端に α_{speed} の速度以上をもって存在した場合，次の時刻に消失するものとして同様にラベルを 0 とする．

次に，各ラベル内で 2 クラスにクラスタリングを行う．各クラスタの中心距離が閾値 α_{center} を上回った場合に，これを分離とみなし，片方に既存ラベルを，もう片方に新規ラベルを与える．

その後，新規出現粒子とラベルを削除されたが残存している粒子に対して，新規ラベルを付与するかどうかの判断を行う．状態空間をメッシュで区切り，各メッシュ内の各該当粒子の尤度の平均が閾値 α_{mesh} を上回っていた場合，そのメッシュ内の該当粒子に対して新規ラベルを付与する．

最後に，各ラベルの中心位置が閾値 β_{center} を下回るほど接近した場合に，それらのラベルを統合する．

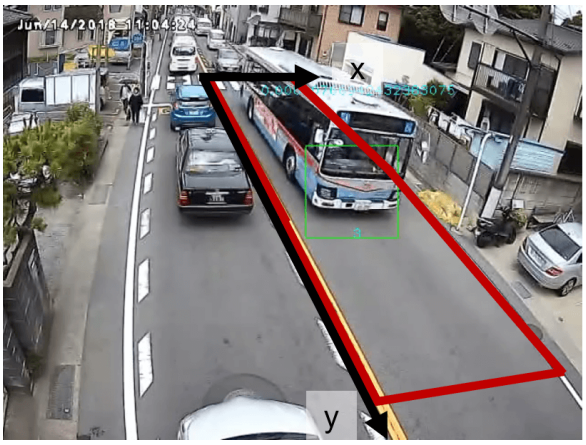
f)　リサンプリング

各粒子を重みに従ってリサンプリングする．本手法では，重みの偏りによる粒子の集中を防ぐため，層化サンプリングを用いた．

5. 実データへの適用と検証

(1)　データ

実データとして，車両追跡には鶴岡八幡宮前交差点の東側を撮影した解像度 640px×480px，10fps の動画像データ（提供：国土交通省関東地方整備局横浜国道事務所）を，歩行者追跡には東急たまプラーザ駅の改札付近を撮影した解像度 612px×428px，10fps の動画像データ（出典：瀬尾ら[18]）を用いた．各データの例を追跡範囲と軸の設定とともに図 **1**, 図 **2** に示す．



図–1　車両追跡用動画像データ例

図–2　歩行者追跡用動画像データ例

(2)　基礎手法の適用結果

a)　車両追跡

1420 フレーム，142 秒間の追跡を行った．検出結果を**表- 3** に示す．また追跡精度の評価を RMSE を用いて行った．この結果を**表- 4** に示す．x 軸・y 軸共に十分な精度が得られた．適用時のパラメータを**表- 5** に示す．

また，検出器を同データに 43 分間適用した結果，実通過台数 211 台対し，検出数 197 台という結果を示した．

b)　歩行者追跡

本手法を歩行者 7 人，計 351 フレームに適用した．本適用では初期分布は手動で与えた．RMSE での精度評価の結果を**表- 6** に示す．適用時のパラメータを**表- 7** に示す．

表–3 車両の検出結果

	実在	存在しない
検出	10	2(共に追跡時排除可)
未検出	2	n/a

表–4 基礎手法の車両適用時における位置・速度精度の RMSE

	画素単位	空間
位置の x 軸方向 RMSE	12.4 px	0.27 m
位置の y 軸方向 RMSE	86.5 px	1.88 m
速度の RMSE	111.4 px/s	8.72 km/h

表–5 基礎手法の車両適用時の各パラメータ

ρ	100
σ_x	10
σ_y	10

表–6 基礎手法の歩行者適用時における位置･速度精度の RMSE

	画素単位	空間
位置の RMSE	49.4 px	1.04 m
速度の RMSE	4.27 px/s	0.32 km/h

表–7 基礎手法の歩行者適用時の各パラメータ

ρ	100
$\boldsymbol{\sigma}$	$0.2\boldsymbol{v}$

(3) 拡張手法の適用結果

目視において混雑率が低い場合では追跡が可能であった．このため追跡を確認できた 6 人に，計 142 フレーム適用した．RMSE での精度評価の結果を**表- 8** に示す．適用時のパラメータを**表- 9** に示す．

(4) 考察

a) 基礎手法の車両適用時に関する考察

本手法では各物体の種類 (クラス) を判別する CNN モデルを検出にも用いている．このためカメラにとらえられた車両の面がどの面であっても検出が可能であることが予測できる．また適用対象エリアは自転車やバイク，歩行者なども通過するが，これらに関しては検出対象として除外できることも予測できる．適用の結果，これらを確認することができた．例を**図 3** に示す．

追跡誤差が大きくなってしまったパターンを大きく 2 つに分け，それぞれについて考察を行う．

表–8 拡張手法の歩行者適用時における位置･速度精度の RMSE

	画素単位	空間
位置の RMSE	22.6 px	0.47 m
速度の RMSE	71.3 px/s	5.39 km/h

表–9 拡張手法の各パラメータ

ρ	70
p_s	0.9
p_D	0.88
σ	$0.3\boldsymbol{v}$
μ_Γ	2
$\beta_{likelihood}$	0.05
$\beta_{position}$	4
α_{speed}	5
$\alpha_{spawned}$	50
α_{mesh}	0.08
β_{center}	40

1 つ目は速度が速い車両追跡時の画面手前領域での誤差であり，**図 4** のような例である．本適用データでは，カメラの設置角度が緩いため，カメラに映る車両の面が大きく変化する．特に手前領域においては，車両の前面よりも側面のほうが大きく写る．ここで観測モデルの性質に着目すると，作成した認識モデルが車両横面に対しても高い尤度を示す性能のために，側面に対応する粒子の尤度も大きくなってしまう性質を持つ．検出時から次の時刻への粒子の速度よりも，車両の速度がある一定以上速い場合，車両の速度についていけず，車両の側面を参照してしまう粒子が多くなる．この場合，上記の性質のため，車両側面にも高い尤度を示してしまい，結果的に実位置よりもやや後ろの位置を推定してしまっていると考えられる．

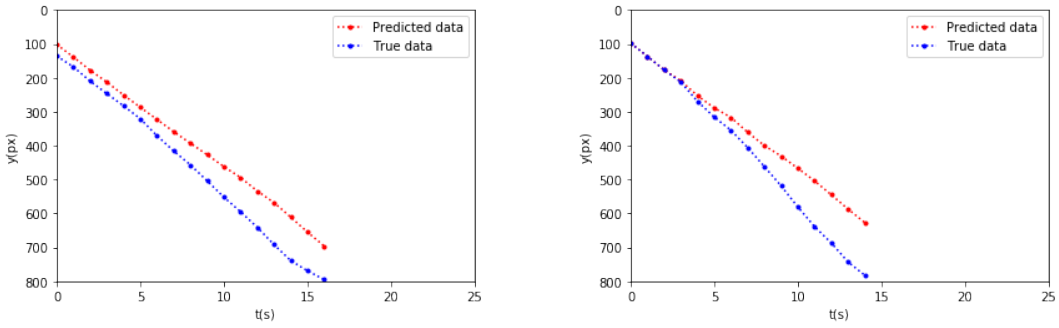
2 つ目は比較的小さい車両追跡時の検出地点付近での誤差であり，**図 5** のような例である．この原因としては，検出時のウィンドウサイズが固定であるため，検出時においては前面の座標ではない部分を参照しうること，また 1 つ目のケース同様，カメラの角度が緩いため，観測空間で数 px ずれるだけでも，状態空間で大きな差を生んでしまうことなどが原因として考えられる．

b) 基礎手法の歩行者適用時に関する考察

第 4 章で述べたように，方向転換時に転換前の方向上に人が連なって存在する場合に大きな誤差が生じた結果となった．この例を**図 6** に，またその追跡時のプロット図を**図 7** に示す． この原因として考えられるものは，**第 4 章**冒頭でも述べたため，ここでは省略する．

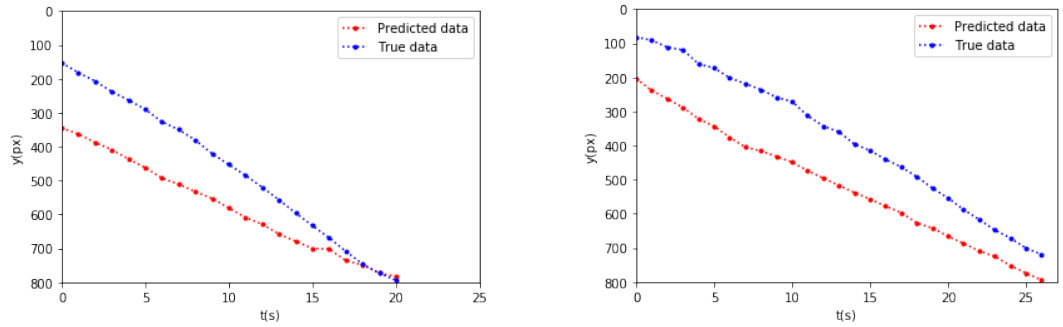
(a) (b)

図–3 見えへの頑健性を示した例 (左)，目的対象以外を検出しなかった例 (右)



(a) (b)

図–4 速度が速い車両追跡時の画面手前領域で大きな誤差が生じたケースのプロット図



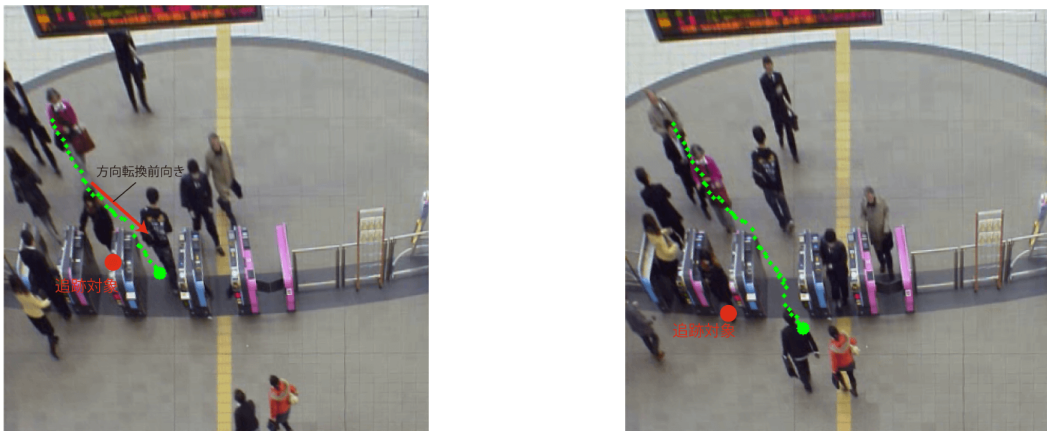
(a) (b)

図–5 比較的小さい車両追跡時の検出地点付近で大きな誤差が生じたケースのプロット図

c) 拡張手法に関する考察

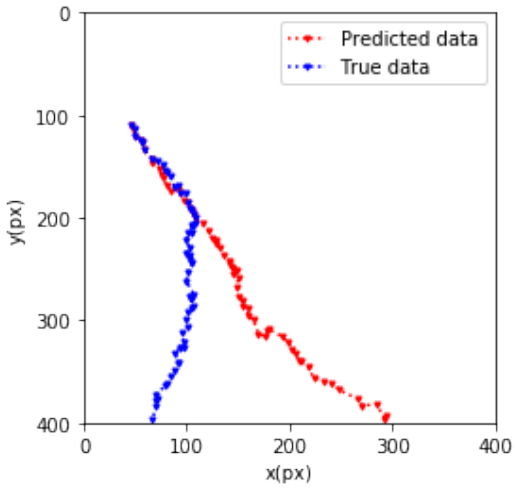
まず追跡可能であった対象に関する結果の例を見てみると，**図 8** のように通った改札までも特定できる精度できているものもあり，位置の RMSE の結果も PF を上回るものとなっている．

拡張手法の適用結果では，そもそも追跡が不可能なケースが存在した．このケースについて考察を行う．**図 9** は追跡が不可能であったケースの 1 例である．緑の点がラベルの中心位置を示しており，その他の点は各粒子を示している．粒子は黒から青に近づく程，尤度が高いことを示している．これを見ると，人物同士の間にも高い尤度を示した粒子が存在していることが分かる．このことから要因の一つとして，適用した CNN モデルが人物の全身のみならず，体の一部分が入力された場合においても全身が入力された際と同程度の尤度を出力可能であるという頑健性により，人物同士が接近した際，個人を識別可能な PHD にならないことが考えられる．



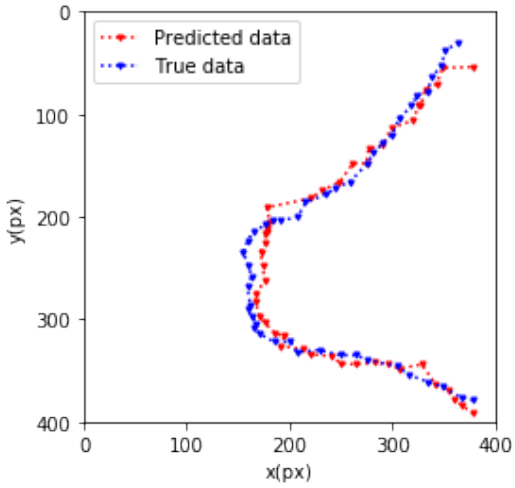
(a) (b)

図–6 基礎手法の歩行者適用時に大きな誤差を生じたケース



図–7 基礎手法の歩行者適用時に大きな誤差を生じたケースのプロット図

図–9 検出不可能なケースの例



図–8 拡張手法による追跡例のプロット図

6. おわりに

本研究では，深層学習の画像認識精度に着目し，パーティクルフィルタの枠組みの中で CNN を観測モデルとして導入した物体追跡のための基礎手法を構築した．そして，この手法に予測される課題を解決するために，PHD フィルタに CNN を導入した拡張手法を構築した．その後，各手法を車両，歩行者用動画像データに適用し，性質を検証した．これにより，基礎手法が車両追跡に対して有効であることを示した．また，拡張手法が歩行者追跡に対して検出可能な場合は有効であることと，追跡不可能であるケースの一要因を示した．今後の課題として，拡張手法の問題点を解決するため，追跡対象のマッチングが可能な CNN の導入や，対象間の影響をシステムモデル内で考慮可能な歩行者挙動モデルなどの導入が有望であると考えられる．

謝辞： 本研究は国土交通省新道路技術会議の研究課題

「学習型モニタリング・交通流動予測に基づく観光渋滞マネジメントについての研究開発」の助成を受けた．PHDフィルタの実装にあたり，東京工業大学の中西航助教に多大な助言を受けた．ここに謝意を表します．

参考文献

1) 布施孝志, 中西航: 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 (土木計画学), Vol. 68, No. 2, pp. 92–104, 2012.
2) ImageNet: Large Scale Visual Recognition Challenge History, http://image-net.org/challenges/LSVRC/2016/index/history.
3) 中西航: 予測モデルと観測データを統合した人物追跡手法の開発, 博士論文，University of Tokyo (東京大学), 2014.
4) 黒畑寿来: 非同期カメラ群による高速道路上における車両軌跡推定, 卒業論文，University of Tokyo (東京大学), 2018.
5) Mahler, R. P.: A theoretical foundation for the stein-winter probability hypothesis density (phd) multitarget tracking approach, Technical Report, ARMY RESEARCH OFFICE ALEXANDRIA VA, 2000.
6) Vo, B.-N., Singh, S., Doucet, A. et al.: Sequential monte carlo implementation of the phd filter for multi-target tracking,Proc. Int'l Conf. on Information Fusion, pp. 792–799, 2003.
7) Ristic, B., Clark, D., and Vo, B.-N.: Improved smc implementation of the phd filter, in *2010 13th International Conference on Information Fusion*, pp. 1–8, IEEE, 2010.
8) Vo, B.-N. and Ma, W.-K.: The gaussian mixture probability hypothesis density filter, *IEEE Transactions on signal processing*, Vol. 54, No. 11, pp. 4091–4104, 2006.
9) Ikoma, N., Haraguchi, Y., and Hasegawa, H.: On an evaluation of tracking performance improvement by smc-phd filter with intensity image of pedestrians detection over on-board camera using neural network, in *2014 World Automation Congress (WAC)*, pp. 273–278, IEEE, 2014.
10) Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
11) Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
12) Ma, C., Huang, J.-B., Yang, X., and Yang, M.-H.: Hierarchical convolutional features for visual tracking, in *Proceedings of the IEEE international conference on computer vision*, pp. 3074–3082, 2015.
13) Cui, Z., Xiao, S., Feng, J., and Yan, S.: Recurrently target-attending tracking, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1449–1458, 2016.
14) 樋口知之編著: 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, 2011.
15) Dalal, N.: INIRIA Person Dataset, http://pascal.inrialpes.fr/data/human/.
16) faezetta: Vehicle Make and Model Recognition Dataset (VMMRdb), https://github.com/faezetta/VMMRdb.
17) 生駒哲一: ランダム有限集合状態空間モデルと逐次モンテカルロフィルタによる動画像中の複数移動物体追跡, 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2010, No. 8, pp. 1–8, 2010.
18) 瀬尾亨, 柳沼秀樹, 福田大輔: Plan-Action 構造を考慮した歩行者挙動モデリングとその適用—駅改札付近を対象として, 土木学会論文集 D3（土木計画学）, Vol. 68, No. 5, pp. I_679–I_690, 2012.

MULTI-OBJECT TRACKING BY DATA ASSIMILATION WITH CONVOLUTIONAL NEURAL NETWORK

Kenta ISHII, Toru SEO and Takashi FUSE