Vol.2021-MBL-101 No.28
Vol.2021-ITS-87 No.28
2021/12/8

複数LiDARによる大規模三次元点群を用いた歩行者トラッキング手法の実装と評価

右京 莉規[1]　天野 辰哉[1]　廣森 聡仁[1]　山口 弘純[1]　守屋 充雄[1,2]

概要：
公共施設や商業施設など様々な人々が行きかう空間における人流計測の需要が高まっている．我々の研究グループではこれまでに，単一の深度センサが捉える三次元点群データを用いて公共空間や準公共空間における歩行者のトラッキング（軌跡導出）を行う手法を提案してきた．同手法ではオクルージョンやノイズによる点群の欠損および複数人物の接近による点群の結合といった観測点群の不完全性による人物セグメンテーションの失敗を考慮し，カルマンフィルタとそれら状況の推定を組み合わせることで，堅牢なトラッキングを実現している．本研究ではトラッキング手法を拡張し，単体の深度センサだけではなく複数の三次元測域センサ（LiDAR）により捉えられた三次元点群の入力に対応し，また実環境においてその有用性検証を行った．都市部繁華街において設置された最大検知距離 260m の LiDAR8 台から得られる広域の三次元点群を約 2 週間継続的に収集したデータを用いて提案手法の評価を行った．その結果，複数人のトラッキングにおける人検出および ID 割り当ての精度を表す MOTA が 0.849 であり，観測セグメントの合体・分裂予測を組み合わせることで，カルマンフィルタベースのトラッキングの精度が 4.7%向上することが確かめられた．

1. はじめに

近年，公共施設や商業施設など様々な人々が行きかう空間における人流検知の需要が高まっている．COVID-19 感染拡大防止のため，ビルや施設の事業者や管理者は人流計測により施設利用者数の把握や密検知を行う必要性に追われている．感染防止のみならず，滞在者数を常時把握しておくことは，地下街などの構内において突発的な火災やゲリラ豪雨等による水害の危険性に対しても効率的な避難誘導につなげることができる．また建築設備設計やスマートビルディングにおける空調サービス最適化などにも活用できるため，利点は大きい．

画像処理技術の飛躍的な発展により，近年では RGB 動画像などから人物を検出するシステムや手法も数多いものの，それらは基本的に顔などの個人情報を含む情報を直接取得するものであることから，利用後ただちに廃棄される場合にも通行者のプライバシーへの不安を払拭することは難しい．また通行者からのオプトアウトの申立てに対応する必要があることや，大多数が納得する十分な説明や同意の取得など多くの作業が必要となることから，公共空間や準公共空間において画像による人流計測を実施することは障壁も多い．これに対し，物体への距離情報のみを取得する三次元測域センサ（LiDAR）や深度カメラなどの三次元距離センサを用いて，より低いプライバシーリスクで人物の存在や姿勢を検出する手法が注目を集めている．三次元距離センサは赤外線パターンの照射とカメラ視差を用いる方法や，赤外線ビームの ToF（Time of Flight）計測により，センサからの各方位に対し最も近い物体への距離を計測し，三次元点群を構成する．

我々は文献 [1] において，単体の三次元深度センサを通行者の側面付近に設置し，取得した通行者の三次元深度データを用いて，歩行者のトラッキング（軌跡導出）を行う手法を提案している．三次元点群を用いる人物検出手法の多くは定位置の人物の姿勢検出を目的としており，歩行者同士が視野を遮る状況は基本的に想定されていない．一方，公共空間における人流検知では人物同士のオクルージョンや太陽光などにより，人物を捉えた点群情報は不完全であることが多い．提案した同手法では，点群の欠損および複数人物の接近による点群の結合といった点群の不完全性による人物セグメントの検出失敗を考慮し，複数歩行者の点群が合体して単一のセグメントとして観測される場合および単一人物の点群が複数のセグメントとして観測される場合の 2 つの状況の推定とカルマンフィルタによるトラッキングを組み合わせることで，堅牢なトラッキングを実現する．

[1] 大阪大学大学院情報科学研究科
[2] 株式会社 HULIX

Vol.2021-MBL-101 No.28
Vol.2021-ITS-87 No.28
2021/12/8

本研究報告では提案手法を拡張し，広範囲の計測が可能な 3 次元 LiDAR を複数台設置することで得られる大規模な三次元点群に対して，歩行者トラッキングを可能にしている．その上で，実環境における広範囲の歩行者トラッキングにおいて提案手法の有用性を示すために，都市部屋外において最大検知距離 260m の LiDAR を 8 台設置し，都市部における広域の三次元点群を約 2 週間継続的に収集し提案手法の評価を行った．その結果、複数人のトラッキングにおける人検出および ID 割り当ての精度を表す MOTA が 0.849 であり，観測セグメントの合体・分裂予測を組み合わせることで，カルマンフィルタベースのトラッキングの精度が 4.7%向上することが確かめられた．

2. 関連研究

複数の歩行者トラッキングを行っている研究には，三次元点群ベースの手法 [2] や Deep SORT[3] などがあげられる．文献 [2] は，三次元点群を入力とし，SECOND[4] とよばれる深層ニューラルネットワークを使用して人物検出を行うとともに，検出した人物セグメントに対して 3D カルマンフィルタとハンガリアンアルゴリズムを使ってトラッキングを行っている．同手法では KITTI データセットに対して評価実験を行なっており，オクルージョンした物体が再出現した時にトラッキングを継続できることを示している．しかし同手法では，観測点群が不完全である場合は考慮されていないため，本稿の手法が想定するような単一のセンサによるセンシングに適用することは難しい．我々の研究グループでも，複数の LiDAR を連結して用いることで，広範囲において人の軌跡を高精度で検出する “ひとなび” システムを開発しており，大型ショッピングモールなどへの設置実験などを実現している [5] が，2 次元データを対象としており，また不完全な点群に対しては堅牢でない．画像に対する最近のトラッキング手法としては Deep SORT[3] がよく知られている．Deep SORT はその前身である SORT[6] の欠点である，オクルージョン後の再出現時に元の ID でトラッキングを行えないといった問題を深層学習を利用して解決している．この手法では，RGB 画像から YOLOv3 を用いて人物検出を行ったものを入力としており，カルマンフィルタを使って前後のフレームで大きさと動きの近いバウンディングボックスを対応させてトラッキングし，KITTI データセットに対して高精度なトラッキングを実現している．

個々の歩行者の観測からの移動推定は Multiple Object Tracking（MOT）とよばれており，カメラ映像を用いたトラッキング技術チャレンジやデータセット共有なども盛んである [7] が，視野が限られる環境における三次元点群データを対象とした試みはほとんどみられていない．なお，既存のトラッキング用のデータセットには MOTChallenge[7] や KITTI[8] が存在する．MOTChallenge[7] は歩行者検出用のデータセットであるが，入力が RGB 画像である．そのため，三次元点群でのトラッキングに利用することはできない．KITTI[8] は通行する車から LiDAR センサを用いて取得した三次元点群のデータであり，一般の通行者を多く含むデータである．しかし，通常の道路や大学構内などで取得されたデータにより構成されており，スーツケースやベビーカーなどを持つ歩行者のデータは含まれない．また背景が一定でないため，固定センサでのデータ処理や評価には向いていない．

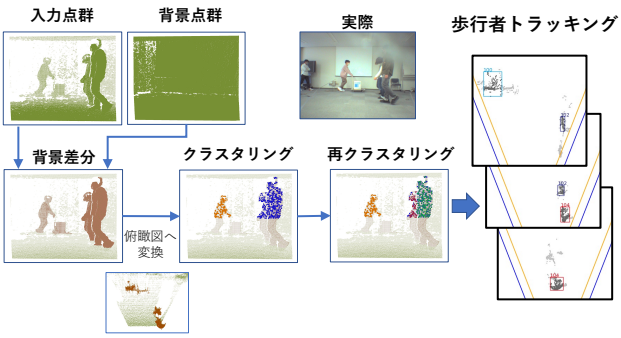


図 1 提案する歩行者検出・トラッキング手法

3. 歩行者セグメントの検出

深度センサや LiDAR などのセンサから得られる点群から歩行者トラッキングを行うため，まずセンサの各フレーム（点群）から歩行者により生じる点群（歩行者セグメント）を検出する．歩行者セグメント検出の流れを**図 1** に示す．

3.1 背景差分による前景点群の抽出

提案手法では，まず背景差分により移動物体のなす点群を含んだ前景点群を抽出する．そのための背景点群として，事前に各センサごとに捕捉可能な領域に移動物体が存在しない時の点群を取得しておいたものを利用する．移動物体が LiDAR により捕捉可能な領域に進入すると，取得された点群と背景点群の間に差分が生じる．その差分の点群を抽出することにより，移動物体の点群（前景点群）を取得する．

3.2 複数センサからの前景点群の統合

複数センサを用いる場合，全センサからの前景点群を統合し，一つの全体の前景点群を生成する．計測する空間の座標系（絶対座標系）内での各センサの 3 次元位置と向きを基にそれぞれの前景点群の座標を絶対座標系上に変換し統合する．統合された点群において複数センサの視野が重複する領域では，同一物体の同一領域に対して重複する点群データが生じうるため，センサ視野が重複しない領域と比較し，得られる前景点群の密度が高くなる．これは後述

する密度ベースの点群クラスタリングに影響を与える．そのため Voxel Grid Filter[9] を使用し，得られた統合前景点群全体に対して各ボクセルグリッド内で重複する複数の点を一つの点にまとめるダウンサンプリングを適用する．ボクセルグリッドのサイズは使用するセンサの計測距離精度から決定する．

3.3 クラスタリングによる歩行者セグメントの抽出

統合された前景点群にクラスタリングを適用し，異なる複数の移動物体のセグメントに分割する．クラスタリング時の処理時間を短縮するために，鉛直方向の軸を取り除き三次元点群を二次元点群に圧縮する．公共空間内に存在する移動物体は二次元平面を移動する人およびその人が所持，利用しているものであるため，鉛直方向の複数人の重複を考慮する必要はないためである．

二次元に変換した点群に対し点群密度に基づくクラスタリングである DBSCAN を適用し，移動物体セグメントを得る．本研究ではこのようにして得られた移動物体のセグメントを歩行者セグメントであるとする．各セグメントは空間座標の軸に沿ったバウンディングボックス（セグメント内の各点の座標の各軸ごとの最小値および最大値の組み合わせ）で表現する．

4. 歩行者のトラッキング手法

検出された歩行者セグメントを観測とし，観測されトラッキングの対象となった歩行者を歩行者オブジェクトと呼ぶ．提案手法では歩行者オブジェクトの状態をカルマンフィルタにより補正・更新・予測しトラッキングする．各オブジェクトはカルマンフィルタの状態として，オブジェクトの位置座標の xy 座標成分，およびオブジェクトの速度ベクトルのうち xy 軸と並行なベクトル成分からなる 4 項組を用いる．

ここでは k 番目のフレームで検出された観測（歩行者）セグメントの集合を $O_k = \{o_{k,j}\}$ とする．j は観測されたセグメントの番号を示す．また k 番目のフレームの処理終了後のトラッキング中のオブジェクトの集合を $T_k = \{t_{k,i}\}$ とする．i はトラッキング中のオブジェクトの番号を示す．T_k 中のオブジェクト $t_{k,i}$ はカルマンフィルタにより補正されたものである．以降の説明において，k 番目のフレームの処理中に T_k に新たに追加されたオブジェクトの速度は xy 方向のどちらも 1.0 とする．T_k の各オブジェクト $t_{k,i}$ から予測した $k+1$ 番目のフレームでのそのオブジェクトの予測値を $p_{k+1,i}$ とし，その集合を P_{k+1} とする．T_0 と P_1，フレーム処理開始時の P_k，T_k はすべて空集合とし，$k \geq 1$ であるとする．

4.1 同一人物判定

k 番目のフレームが得られた際の処理について述べる．フレーム処理開始時にあらかじめ T_{k-1} をもとに現在のフレームでのオブジェクトの予測値 P_k を求めておく．P_k が空の場合，$T_k \leftarrow S_k$，つまり現フレームで観測されたセグメントすべてをトラッキングの対象としてこのフレームの処理を終了する．

T_{k-1} が空でない場合，以下の 3 つの処理を行う．

(1) 現フレームのセグメント集合と前フレームでのオブジェクト集合の対応付けを，距離と体積に基づくコストを最小化するように決定する．
(2) (1) で対応付けが与えられなかったオブジェクトおよびセグメントに対し，その原因を推定し，それに応じた対応付けを決定する．
(3) (1) と (2) で決定された対応付けの情報を用い，カルマンゲインと状態更新（カルマンフィルタ処理）を行う．

(1) の対応付けについては，与えられたセグメントとオブジェクトの組に対し，セグメントに含まれる点をすべて内包し，各辺が xyz 軸のいずれかと平行な直方体（バウンディングボックスとよぶ）の体積，およびオブジェクトとセグメントの組から計算される「コスト」の和が最小になるように求める．ここで，現フレーム k のセグメント数（$|O^k|$ とする）とフレーム $k-1$ のオブジェクト数（$|T^{k-1}|$ とする）に対し，以下の条件を満足する対応付けを発見する．

- どのセグメントも高々 1 つのオブジェクトに対応付けられる
- どのオブジェクトも高々 1 つのセグメントに対応付けられる
- $\min\{|O^k|, |T^{k-1}|\}$ だけの対応が存在する
- 対応付けのコスト総和が最小である

これにより，オブジェクトあるいはセグメントの数の少ないほうがすべて対応付けを持つようにする．

(2) では，(3) でのカルマンフィルタの更新方法を決定するために，対応付けの存在しないオブジェクトあるいはセグメントが生じた原因を推定する．オブジェクトやセグメントはオクルージョン（遮蔽）や領域外への移動による消失や，領域内への移動による出現が発生するが，それらの発生原因の領域内の位置に応じた推定を行うことで (3) のカルマンフィルタの精度向上を図ることを目的とする．

最後に (3) では，得られた対応付けに基づき，カルマンフィルタを用いてフレーム k でのオブジェクトを予測する．カルマンフィルタの状態として，オブジェクトの位置座標の xy 座標成分，およびオブジェクトの速度ベクトルのうち xy 軸と並行なベクトル成分からなる 4 項組を用いる．また観測は，現フレームのセグメントを内包するバウンディングボックスの中心座標成分とセグメントの 3 辺の大きさからなる 6 項組で表す．これらを用い，例えばあるオブジェクトに対し対応するセグメントが見つかっていない場合でも，それがオクルージョンによるセグメント消失

と想定される場合にはトラッキングを継続する．(2) での対応づけが存在しないオブジェクトについては，観測値なしでカルマンフィルタを更新する．T_k での速度でセグメントが移動したと仮定しカルマンフィルタを更新する．対応付けが存在するオブジェクトについてはカルマンフィルタを用いて現在の位置を更新する．

(1), (2) の処理について，以下の 4.2 節から 4.3 節で詳細を述べる．

4.2 オブジェクトとセグメントの対応付け

前節で述べたように，前フレームでのオブジェクトと，現フレームでのセグメントの各組に対し，それらの距離をその組の「コスト」とし，前節で述べたような条件を満たしながらコストを最小とする対応付けを発見する．オブジェクト i とセグメント j に対するコスト $c_{i,j}$ は以下で定義される．

$$c_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (1)$$

x_j, y_j はそれぞれ観測セグメント $o_{k,j}$ の xy 座標を表し，x_i, y_i はそれぞれ歩行者セグメントの予測値 $p_{k,i}$ の xy 座標を表す．このコスト $c_{i,j}$ が最小となる組み合わせから順に割り当てを決定する．なお，決定された組み合わせのコストが閾値（経験的に 1.0 を利用）を超える場合，その割り当てを棄却する．

割り当てられた，観測と P_k 中の歩行者の組については，その観測セグメントとカルマンフィルタを用いて，対応する T_{k-1} 中の歩行者の現在位置を更新し，T_k に追加する．

4.3 対応付けされなかったオブジェクトあるいはセグメントの処理

前節の方針に基づいて決定された対応付けに含まれないオブジェクトあるいはセグメントが存在する場合，本節の追加対応付け手法を適用し，それらの対応付けを発見する．同手法は，対応付けがなされなかった原因を推定し，その理由に応じ，最も適切と思われる対応付けを提供する．

あるセグメントが対応するオブジェクトを持たなかった原因として，本研究では，以下の 3 つであると考える (**図 2**)．

(S1) 新しい歩行者が領域に進入したためにセグメントを得るが，その前フレームに（それと対応付け可能な）オブジェクトが存在しない．例えば図 2 の S1 では，領域外の歩行者（白で表示）がセンサー可視領域に進入したために新たにセンサーに捉えられ，赤いバウンディングボックスで示されるセグメントが得られた様子を示している．

(S2) オクルージョンの影響で，1 人の歩行者に対し l 個に分裂したセグメントを得るが，その前フレームに，それらに対応するオブジェクトは 1 つしか存在しない（すなわち残りの $l-1$ 個のセグメントに対する対応付けが得られない）．例えば図 2 の S2 では，ストローラーを押す歩行者（青いバウンディングボックスで表示）のオブジェクトに対し，その次フレームではセンサー付近の歩行者のオクルージョンにより，左側と右側の 2 つの赤いバウンディングボックスで示される 2 つのセグメントが得られた様子を示している．

(S3) オクルージョンの影響で観測されていなかった歩行者が再度観測されたためにセグメントを得るが，その前フレームに（それと対応付け可能な）オブジェクトが存在しない． 例えば図 2 の S3 では，前フレームではセンサー付近の歩行者のオクルージョンにより，オブジェクトが存在しなかったが，現フレームではオクルージョンがなく，赤いバウンディングボックスで示される 1 つのセグメントが得られた様子を示している．

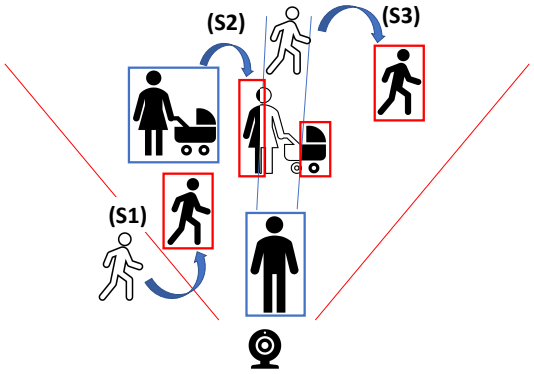


図 2 セグメントに対応するオブジェクトが存在しない原因

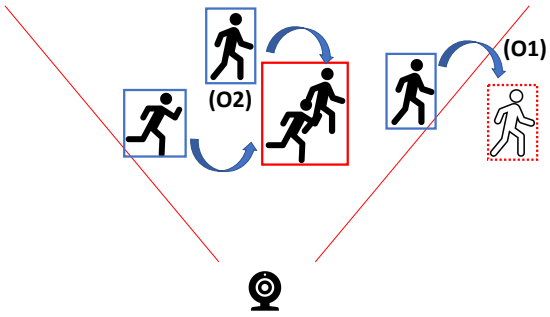


図 3 オブジェクトに対応するセグメントが存在しない原因

また，あるオブジェクトが対応するセグメントを持たなかった原因として，本研究では以下の 2 つであると考える (**図 3**)．

(O1) 前フレームでオブジェクトとして認識されていた歩行者が領域外に退出し，対応するセグメントを失った結果，対応付けが得られない（図 3 の O1）．

(O2) 複数人が接近して通行したことにより，2 人以上の歩行者に対し 1 つのセグメントしか得られない（図 3 の

O2)．

これに対し，(S1) および (S3) の場合は，新たに歩行者が出現したとみなしてよいため，次フレーム以降でカルマンフィルタによる位置予測を開始すればよい．したがって，このセグメントに対するカルマンフィルタは適用せず，これを現フレームでのカルマンの状態（オブジェクト）として追加する．(S2) の場合は，前フレームで 1 つのオブジェクト i として認識されていたものが，現フレームではオクルージョンにより l 個のセグメントとして観測された場合（観測が「分裂」した場合）である．これが理由と想定される場合，このオブジェクト i のトラッキングをこの分裂発生時で終了し，以降は分裂により発生したセグメントをそれぞれトラッキングする．

(O1) の場合は，領域外へ歩行者が退出したとみなしてよいため，このセグメントに対するカルマンフィルタは適用しない．また，F_{thr} フレーム以上連続してセグメントとの対応付けがないオブジェクトはトラッキングを終了する．(O2) の場合は，前フレームで n 個のオブジェクト i_l $(l = 1, 2, 3, ..., n)$ として認識されていたものが，現フレームではオクルージョンにより 1 個のセグメントとして観測された場合（観測が「合体」した場合）である．これが理由と想定される場合，各オブジェクト i_l の予測値 p_{k,i_l} と観測 $o_{k,j}$ の重なり合う領域を仮想セグメントとして生成し，この仮想セグメントを i_l のセグメントとして対応付ける．

5. 評価実験

固定設置した複数の 3 次元 LiDAR から 3 次元点群を収集し，提案手法による広域の歩行者トラッキング性能を評価した．

5.1 実験環境および収集データ

実験環境を図 4 をおよび図 5 に示す．都市部の約 110m × 40m の範囲にわたる屋外領域において施設管理者の許可を得て設置された 3 次元 LiDAR を 8 台から，歩行者を含む空間の 3 次元点群を収集した．3 次元 LiDAR として LIVOX 社の Livox Horizon を使用した．表 1 に本 LiDAR の仕様を示す．各 LiDAR は図 5 のように，道路沿い街灯の，地面からの高さが約 3.5m の位置に固定した．初夏の 2 週間にわたり，毎日午前 10 時から午後 8 時の間収集したデータを用いた．

3 節で述べたように，各 LiDAR から得られるそれぞれの点群に対して，事前に取得した背景点群データを用いて背景差分を適用し，前景点群を毎フレーム取得した．さらに 8 台の LiDAR の相対位置を基にそれぞれの LiDAR からの前景点群を統合し，空間全体の統合された前景点群を得た．旗や樹木の葉，ロープなど頻繁に動くために背景差分処理によって前景点群，つまり移動物体として捉えられる物体に対しては，事前にその物体を含む領域を手動で指定し，指定領域内の点群すべてを統合前景点群から除外することで対応した．周辺建造物の内部の点群も窓ガラスや建物入り口を通して前景点群として取得されることがあったため，同様に建物領域の点群を統合点群からすべて除去した．統合および除去後の前景点群は 1 フレームあたり平均 18,984.5 点の点群を含んでいた．この点群に対して歩行者セグメント検出およびトラッキングを適用しその性能を評価した．

表 1 3 次元 LiDAR Livox Horizon の性能

項目	性能
点群数	480,000 点/秒
フレームレート	10 フレーム/秒
最大計測可能距離	260 m
走査視野角（水平）	81.7 度
走査視野角（垂直）	25.1 度
距離精度	± 2.0 cm
角度精度	± 0.05 度

5.2 評価指標

複数人のトラッキング性能を評価するため，本研究では文献 [10] で導入された CLEAR 指標のうち以下の評価指標を採用した．

- MOTA: 複数人のトラッキングにおける人検出および ID 割当ての精度を表す．式 (2) にて計算される．

$$MOTA = 1 - \frac{\sum_t (FP_t + FN_t + IDsw_t)}{\sum_t GT_t} \quad (2)$$

  式 (2) における FP_t，FN_t，$IDsw$，GT_t は，フレーム t での ID 割当ての誤り数（ただし $IDsw$ に相当する場合は除く），検出の失敗（見逃し）数，ID 割当てにおける ID 入替わり数，およびフレーム t での真の人数をそれぞれ表す．なお，$FP_t + FN_t + IDsw_t$ の最大値は GT_t である．
- MT: 全歩行者のうち，80%以上の期間で正しくトラッキングを行えた歩行者の割合を表す．
- ML: 全歩行者のうち，50%以下の期間でしか正しくトラッキングを行えなかった歩行者の割合を表す．
- FP: 全歩行者・全期間を通じた誤 ID 割当て数を表す．
- FN: 全歩行者・全期間を通じた見逃し数を表す．
- IDsw: 全歩行者・全期間を通じた ID 割当てにおける ID 入替り数を表す．
- Frag: 全歩行者・全期間を通じ，FN に相当する見逃しにより軌跡が分割された数を表す．

5.3 評価結果

データセットのうち約 2 分間のデータに対し，観測セグメントの合体・分裂予測を行う場合（提案手法），および行わない場合の 2 種類で精度評価を行った．その結果を**表 2** に示す．

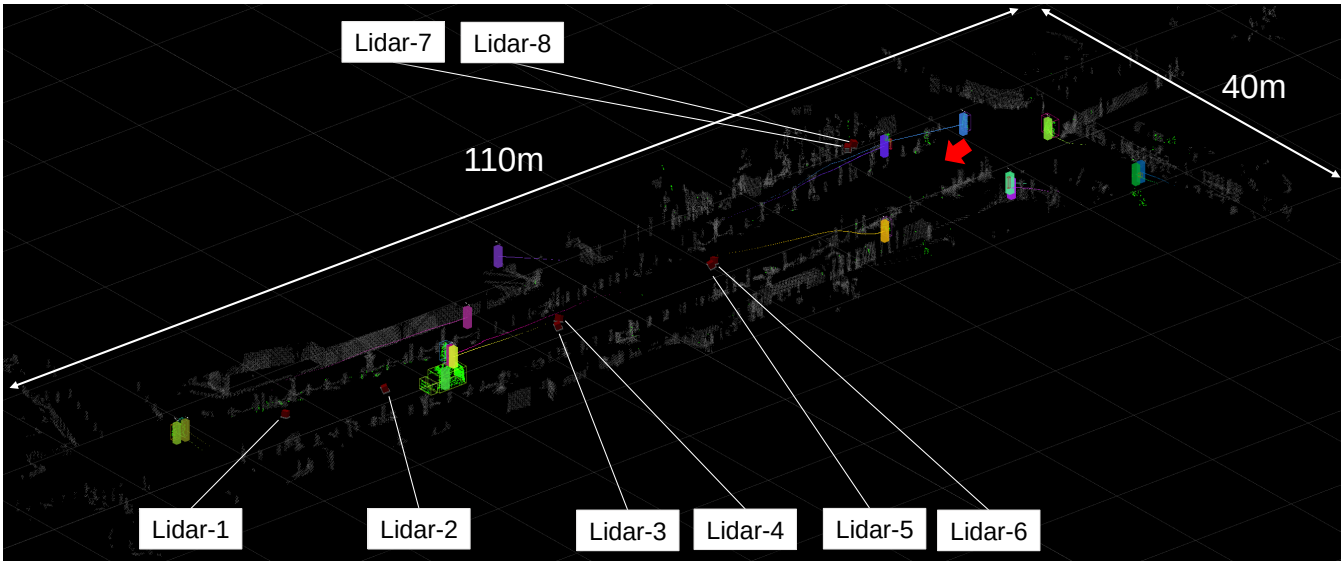


図 4 3 次元 LiDAR 配置および取得される 3 次元点群の斜視図．白色および緑色の点群はそれぞれ取得された背景点群（地面によるもの以外），前景点群を示す．

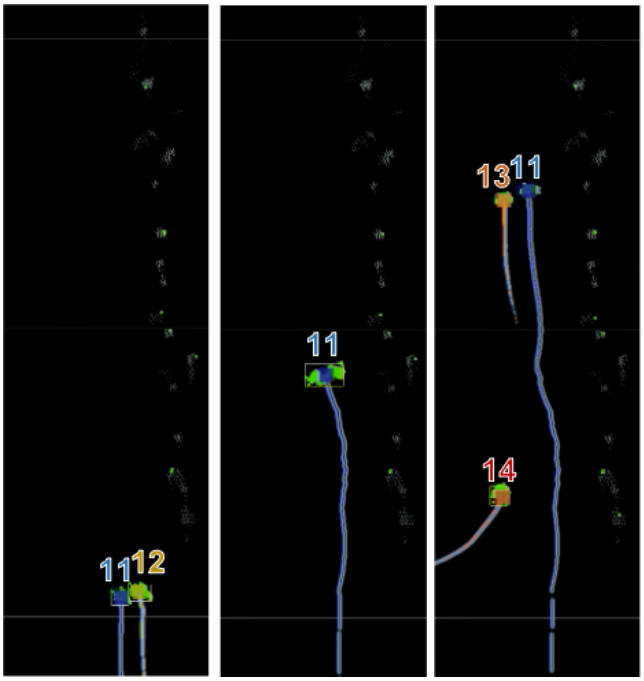
表 2 評価結果

	MOTA	MT(%)	ML(%)	FP	FN	IDsw	Frag
観測セグメントの合成・分裂予測なし	0.802	42.3	34.6	0	85	12	12
観測セグメントの合成・分裂予測あり（提案手法）	0.849	65.4	19.2	0	65	8	10

図 5 3 次元 LiDAR 設置状況および実験環境．図 4 の赤矢印地点から矢印方向に撮影．

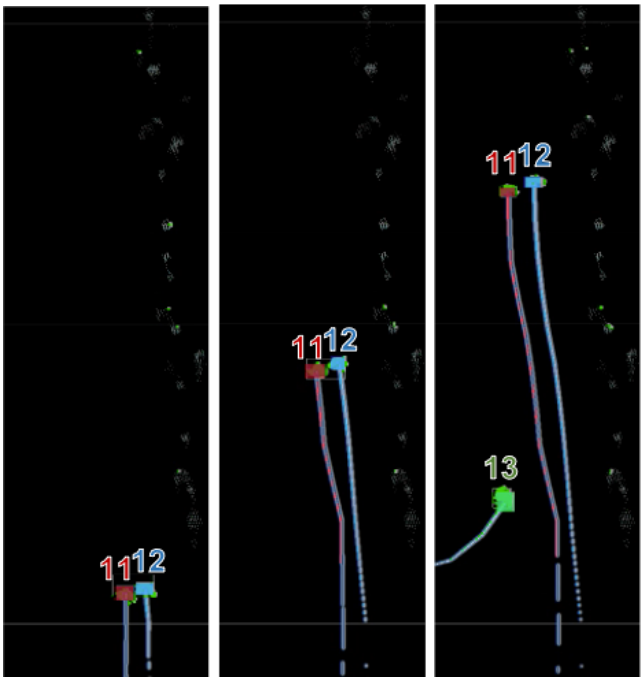
観測セグメントの合体・分裂予測を行うことより，MOTA が 4.7%程度向上していることがわかる．その原因として，FN の値が大幅に減少したことが挙げられる．これは，複数人が接近した際に，複数人を 1 人であると判定された場合のトラッキングの違いによるものである．予測を行わない場合，複数人のうち判定されなかった歩行者をそれ以降のフレームで同一人物であると判定できなかった．それに伴い，判定されていない間は観測セグメントが存在しないことによりトラッキングを継続できず，FN の値を増加させる結果となった．また再び複数人の観測セグメントを検知できた際も，新たな別の歩行者であると判定され，IDsw の値を増加させることとなった．一方，提案手法では，複数人を 1 人であると判定された場合も，観測セグメントを予測している歩行者に分割することにより，観測セグメントが 1 つしか存在しない場合もそれぞれの歩行者のトラッキングを継続することができ，それにより FN の値や IDsw の値を減少させた．

複数人を表す観測セグメントが合体し 1 つになっている例のうち，合体・分裂予測を行わなかった場合の結果を**図 6** に，予測を行った場合の結果を**図 7** に示す．これらの図では，この状況を俯瞰図方向から表示したものであり，同一人物としてトラッキングしているセグメントに対して色付きのバウンディングボックスで示し，またその人物の軌跡を同色で示している．この場面は，(a) で 2 人の歩行者が図の下方から上方に向かい移動しており，(b) で 2 人の歩行者の観測セグメントが 1 つとして検知される．そして，(c) で再び 2 人の観測セグメントが検知される．なお，(c) では別の歩行者が中央付近を通過している．予測を行わなかった場合の結果からは，(a) で観測されていた歩行者 11 と歩行者 12 が，(b) で観測セグメントが合体し歩行者 11 として検知された．この時，歩行者 12 のトラッキングが途絶えた．再び (c) で 2 人のセグメントが検出されると，(a) での歩行者 11 は歩行者 13 となり，また歩行者 12 は歩行者 11 であると，別人物としてトラッキングが行われたことがわかる．一方，予測を行なった場合の結果からは，(b) で観測セグメントが合体した際も 2 人の歩行者の位置予測を継続することができ，(c) で 2 人分のセグメントとなった際にも (a) の歩行者と同一人物として判定しトラッキングを継続できていることがわかる．



(a) (b) (c)

図 6 合体・分裂予測を行わなかった場合のトラッキング



(a) (b) (c)

図 7 合体・分裂予測を行なった場合のトラッキング

6. まとめ

本研究では，都市繁華街において 3 次元 LiDAR を複数設置し，一般の歩行者を含む空間の 3 次元点群を収集した．収集した 3 次元点群からそれぞれの歩行者についてカルマンフィルタを用いて位置と速度を予測することにより，トラッキングを行う手法を提案した．提案手法では，複数の LiDAR から取得した 3 次元点群を，背景差分により前景点群を抽出しそれぞれの LiDAR から抽出した前景点群を統合した．統合した前景点群に対し，クラスタリングにより歩行者を表す 3 次元点群のセグメントを抽出した．複数人の歩行者を表すセグメントの合体・分裂を予測しながらそれぞれの歩行者の位置と速度を予測し，観測された歩行者との割り当てを行うことによりトラッキングを行なった．収集した実際の通行データのうち約 2 分のデータを用いた精度評価の結果，MOTA が 0.849 を達成した．これにより，観測が合体・分裂するような公共空間でも歩行者トラッキングを十分な精度で実現できることを示した．

謝辞 本研究成果は JST CREST「基礎理論とシステム基盤技術の融合による Society 5.0 のための基盤ソフトウェアの創出」の助成を受けたものです．

参考文献

[1] R. Ukyoh, A.Hiromori, H.Yamaguchi, and T.Higashino. 公共空間における三次元点群の不完全性に対して堅牢な歩行者トラッキング手法. 第 185 回 マルチメディア通信と分散処理研究発表会 (DPS 研究発表会), 2020.

[2] B. Sahin, D. Wang, C. Huang, Y. Wang, Y. Deng, and H. Li. A 3D Multiobject Tracking Algorithm of Point Cloud Based on Deep Learning. *Mathematical Problems in Engineering*, Vol. 2020, p. 8895696, 2020.

[3] N. Wojke, A. Bewley, and D. Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, 2017.

[4] Y. Yan, Y. Mao, and B. Li. SECOND: Sparsely Embedded Convolutional Detection. *Sensors*, Vol. 18, No. 10, 2018.

[5] H.Yamaguchi, A.Hiromori, and T.Higashino. A Human Tracking and Sensing Platform for Enabling Smart City Applications. In *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking*, Workshops ICDCN '18, pp. 13:1–13:6, New York, NY, USA, 2018. ACM.

[6] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple Online and Realtime Tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, 2016.

[7] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé. MOT20: A benchmark for multi object tracking in crowded scenes. *arXiv:2003.09003[cs]*, March 2020. arXiv: 2003.09003.

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous Driving? The KITTI Vision Benchmark Suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, 2012.

[9] Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4, 2011.

[10] K. Bernardin and R. Stiefelhagen. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP Journal on Image and Video Processing*, Vol. 2008, No. 1, p. 246309, 2008.