Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

多様な人物属性を考慮した 堅牢な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]　東野 輝夫[1]

概要：
我々の研究グループではこれまでに，荷物やベビーカーなどを持って移動する人々などを含む多様な属性の人物が行きかう通路などにおいて，それらの人物を単一の深度センサから捉えた三次元点群データから歩行者のトラッキングを行う手法を提案している．単一のセンサであればセンサ設置位置の制約も少なく，トラッキングが容易になる一方，センサ付近を通過する人物が視野を大きく遮蔽することにより広い領域でオクルージョンが発生し，その結果，人物に相当する点群のセグメント化が適切に行われずトラッキングが不安定になる傾向があった．本研究では，三次元点群から人に相当する点群をセグメント化する際に，高さ情報を積極的に活用したクラスタリングを行うことで，セグメント化の性能を向上させるとともに，オクルージョンによる点群の欠落を推定・補完することで，カルマンフィルタ適用時における誤差の発生を抑制する手法を提案する．実験環境および実商業施設において市販の小型深度センサを用いて収集した様々な属性の歩行者を含む三次元点群データを用い，提案手法を評価した結果，複数オブジェクトのトラッキング精度指標 MOTA が 0.914 であり，単体センサの狭小な観測範囲内でオクルージョンが多発する環境でも十分な精度でのトラッキングを達成できた．

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 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，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いる．同手法では，取得した三次元深度データを三次元点群データに変換し，背景差分法を適用することによりベビーカーやキャスターを押す人々を含む様々な属性の通行者の点群のみを抽出する．次に，クラスタリングに基づくノイズ除去および個々の通行者に対応するセグメント抽出を行う．それらのセグメントがおおよそ一定速で移動すると仮定し，観測された点群と合わせて新しい位置を予測するカルマンフィルタを複数のセグメントに同時適用することにより通行者の軌跡導出を行っている．

同手法では，三次元深度センサは通路脇などの壁面に設置される一般的な状況を想定しているため，センサ座標系

[1] 大阪大学大学院情報科学研究科

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

の前方（奥行き）方向が通行者の主な移動方向と直交することが多い．したがって，センサ近傍を通過する通行者によるオクルージョンが発生し，他の通行者の全部または一部の点群が観測できない場合が多く発生する．特に，ベビーカーなどを押す人は一般通行者よりも大きなセグメントとして検出されるが，オクルージョンが発生するとその一部が遮蔽されることによりセグメントが矮小化し，一般歩行者と誤って認識されることも少なくない．こういったいわゆる「欠落した観測データ」がカルマンフィルタの予測精度を低下させるといった問題があった．本研究では，三次元点群から人に相当する点群をセグメント化する際に，高さ情報を積極的に活用したクラスタリングを行うことで，セグメント化の性能を向上させるとともに，オクルージョンによる点群の欠落を検出し，カルマンフィルタ適用時における誤差の発生を抑制する手法を提案している．

実験環境および実商業施設において市販の小型深度センサを用いて収集した様々な属性の歩行者を含む三次元深度データを収集した．精度検証では，文献 [2] で導入された CLEAR 指標を利用して評価を行なった．CLEAR 指標では，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 と，複数人のトラッキングにおける位置予測の精度を表す MOTP が指標として用いられる．収集した深度データを用いた精度検証の結果，MOTA が 0.914，MOTP は 0.520 であった．これにより，不完全な三次元点群を用いた場合でも公共空間での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．

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 [3] や Deep SORT[4] などがあげられる．文献 [3] は，三次元点群を入力とし，SECOND[5] 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 [6] 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[4] がよく知られている．Deep SORT はその前身である SORT[7] 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．

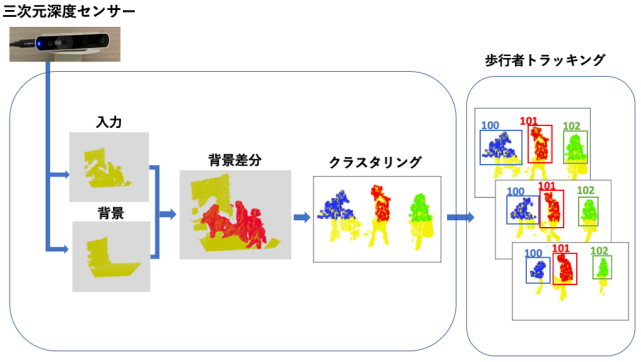


図 1 提案する歩行者検出・トラッキング手法

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 [8] 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[8] や KITTI[9] が存在する．MOTChallenge[8] 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[9] 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．

3. 歩行者の検出手法

単一の三次元深度センサから得られる毎フレームの三次元の深度データを三次元の点群に変換し，前景点群の抽出とクラスタリングによりフレーム内の歩行者により生じる点群（歩行者セグメント）を検出する．提案手法の概要を **図 1** に示す．

3.1 背景差分による移動物体の抽出

提案手法では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．まず背景データとして，三次元深度センサで捕捉可能な領域に移動物体が存在しない時の三次元深度データを取得する．深度センサが取得する三次元

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

深度データには，センサ機器に依存した欠損値がフレームごとにランダムに含まれる．このため，背景データの生成のために複数フレームの三次元深度データを取得し，欠損値を除いた各ピクセルの最頻値を求めることでこの欠損値を除去する．

移動物体がセンサの捕捉可能な領域に進入すると，取得される三次元深度データと背景データ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，移動物体の三次元深度データ（前景点群）を取得する．実際に取得した三次元深度データにはデバイスの影響等により多少の誤差が含まれるため，本研究では観測された三次元空間上の点と背景点群の差が 10cm 以上ある領域のみを前景点群として判定する．

さらにセンサの測定誤差により前景点群に生じるノイズの除去を行う．ここでは前景点群において，センサ設置位置との点との距離がセンサの最大検出距離以上である点を削除する．次に，以降の処理時間短縮のため，抽出した点群データのうち 1024 ポイントをランダムに選択しダウンサンプリングを行う．この際，選択前の時点での点群データのポイント数が 1024 ポイントよりも少ない場合は，そのフレームのデータは歩行者の存在しないフレームのデータとして棄却する．

3.2 クラスタリングによる歩行者セグメントの抽出

背景差分により抽出し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．公共施設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．またクラスタリングの際に，センサよりも高い位置に存在する点群のみを抽出する．これは，複数歩行者が接近した際にセグメント分割の精度が低下しやすく，接近しやすい手や足を除くことで分割精度を向上させるためである．

二次元に変換した点群に対し DBSCAN クラスタリングを適用し，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．

歩行者同士の接近により，複数歩行者が一つのセグメントとして誤検出される場合がある．そのような誤検出を防ぐために，それぞれのセグメント内の三次元点群に対して高さのピークを検出し，各セグメント内に含まれる歩行者数を推定する．

図 3 のように xyz 軸を定義すると，xz 平面上で y 軸方向の大きさを判定し，極大値となる点を検出する．本研究では，歩行者の xz 平面上での大きさを考慮し，5cm 四方の正方形に区切り，それぞれの正方形に含まれる点群のうち y 軸方向の成分の最大値を求め，それぞれの正方形での y 軸方向成分の最大値から極大値を算出する (**図 2**)．極大値が 2 つ以上検出されたセグメントについては，探索半径を小さく最小近傍点数を大きくし再度クラスタリングを適用することで，歩行者セグメントをさらに分離する．

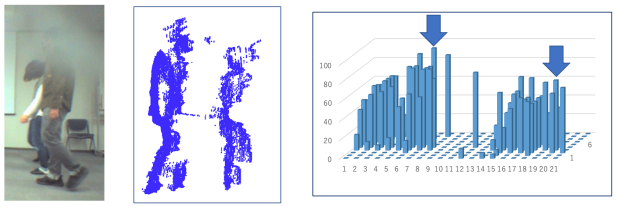


図 2 高さのピーク検出の図．5cm 四方に区切られた正方形それぞれに含まれる点群のうち y 軸方向成分の最大値が右図に示されている．矢印の 2 箇所に極大値が算出されている．

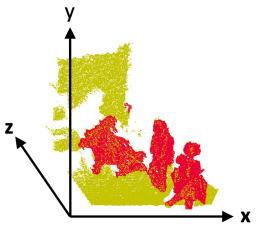


図 3 xyz 軸の定義

4. 歩行者のトラッキング手法

提案手法では歩行者セグメン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．各セグメントは xyz 座標および xyz 方向の速度の 6 つの状態の値を持つ．セグメントの位置を表す xyz 座標は，セグメントの中心座標である．

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．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歩行者セグメントの集合を $T_k = \{t_{k,i}\}$ とする．i はトラッキング中の歩行者の番号を示す．T_k 中の歩行者セグメント $t_{k,i}$ 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歩行者セグメントの速度は xyz 方向の全て 0 とする．T_k の各セグメント $t_{k,i}$ から予測した $k+1$ 番目のフレームでのその歩行者セグメントの予測値を $p_{k+1,i}$ 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歩行者セグメントの予測値 P_k を求めておく．

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

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，まず前フレームからの現在フレームでの歩行者セグメントの予測値の集合 P_{k-1} と今回の観測セグメント集合 O_k の対応付けを求める．P_k と O_k のセグメントのすべての組み合わせに対して，セグメントの位置と体積を用いて (1)〜(3) 式にて求められるコスト c を計算し，最小となる組み合わせから順に割り当てることで決定する．

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

$$V_{i,j} = |V_j - V_i| \quad (2)$$

$$c_{i,j} = \alpha d_{i,j} + (1 - \alpha) V_{i,j} \quad (3)$$

この式での $\alpha = 0.82$ は，位置情報と体積情報の利用割合を表す．$d_{i,j}$ は歩行者セグメントの予測値 $p_{k,i}$ と観測セグメント $o_{k,j}$ の距離を表す．同様に $V_{i,j}$ は歩行者 $p_{k,i}$ と観測値 $o_{k,j}$ の体積の差を表す．x_j, y_j, z_j, V_j はそれぞれ観測セグメント $o_{k,j}$ の xyz 座標および体積を表し，x_i, y_i, z_i, V_i はそれぞれ歩行者セグメントの予測値 $p_{k,i}$ の xyz 座標および体積を表す．なお，観測集合の要素数とトラッキング中の歩行者数が異なる場合は，少ない方の数だけ割り当てを行う．

割り当てられた，観測と P_k 中の歩行者の組ついては，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．カルマンフィルタの更新時，割り当てられている観測セグメントの体積の絶対値およびその歩行者の直前の 5 フレームに対する体積の相対値を元にカルマンゲインを変更する．詳細は 4.5 章で述べる．

4.2 オクルージョンによる観測欠損の補完

公共空間では複数人が同時にセンサの前を通過することがあるため，ある歩行者によって発生する死角となる範囲を別の歩行者が通行することで，観測セグメントが欠損するオクルージョンが発生することがある．このオクルージョンに対応するため，オクルージョンしている観測に歩行者が割り当てられた場合，**図 4** のように観測の補完を行う．観測の補完では，その観測の x 軸および y 軸の長さを，その歩行者の過去の x 軸および y 軸方向の長さの最大値に変更し，それに対応するようにその観測を x 軸の正の方向と負の方向のうちのオクルージョンをしている方向に移動させる．x 軸正の方向と負の方向の両方にオクルージョンしている際は，移動は行わず観測の長さのみ変更する．その後，割り当てた観測と予測される歩行者の位置や体積が大きく異なる場合，その割り当てを棄却する．

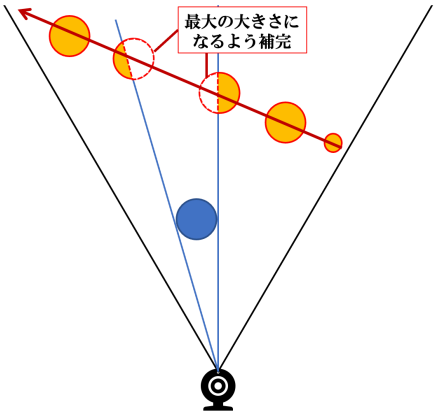


図 4 オクルージョン時の予測．青の歩行者によって橙の歩行者が一部オクルージョンする．橙の歩行者の観測がオクルージョンしている場合，以前のフレームでの最大の大きさになるよう補完する．

4.3 割り当てられないセグメントの処理

予測歩行者セグメント P_k と観測セグメント O_k の数が異なる場合，歩行者セグメントの予測値に割り当てられない観測，あるいは観測に割り当てられない予測値が生じる．この原因は次の 4 種類に分けられる：(1) 新たな歩行者がセンサの画角外から画角内に移動し観測されたため．(2) トラッキングされていた歩行者が画角内から画角外へ移動し観測されなくなったため．(3) 複数人の通行によりオクルージョンが発生し，1 人の歩行者に対し 2 つ以上のセグメントに分裂したため．(4) 複数人の接近により 2 人以上の歩行者のセグメントが 1 つに合体したため．これらの判定方法については 4.4 章にて詳しく述べる．

提案手法では上記の原因に基づいて，未割当の観測や予測値について次のように処理を行う．(1) 新しく画角内に出現した歩行者セグメントは新たにトラッキング中の歩行者の集合 T_k に追加する．4 フレーム連続で (2) 観測に割り当てられず，かつほかの歩行者によるオクルージョン領域に存在しないとみなせる歩行者セグメントは，画角外へ消失したとし，T_k に追加せず，トラッキングから外す．(3) や (4) ケースで分裂や合体により発生したセグメントは，複数人を含む場合や 1 人に満たない場合についても，分裂や合体のないセグメント同様に T_k に追加する．分裂の場合は，分裂を統合したセグメントを，合体の場合には該当する予測値 P_k 中のセグメントを観測値としてカルマンフィルタの更新に利用する．分裂や合体が終了し元の歩行者に対応する観測値を 4 フレーム以上連続で取得すると，分裂や合体によりトラッキングに追加したセグメントのトラッキングを終了する．なお分裂や合体の発生中は，(2) のケースで分裂や合体元の歩行者に対応する観測値が 4 フレーム以上連続で存在しない場合でもトラッキングを継続する．

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

4.4 割り当ての存在しないセグメントの原因推定

本章では，割り当ての存在しないセグメントが発生した場合についての原因を推定する処理について述べる．割り当ての存在しないセグメントが発生する際，歩行者に対応する観測が存在しない場合と観測に対応する歩行者が存在しない場合の 2 種類が存在する．

まず，以前のフレームで予測された歩行者に対応する観測が存在しない場合について述べる．歩行者 i_1 から派生した分裂や合体が存在しているかを推定する．存在する場合，歩行者 i_1 に対応する観測値が存在しない原因は分裂や合体が発生していることであると判定する．存在していない場合，次の推定を行う．次に，歩行者 i_1 の予測位置 $x_{i_1,k-1}$ が観測値 z_j の内部に包含されているかを推定する．包含されていない場合は，歩行者 i_1 に対応する観測値が存在しない原因は画角内から画角外へ移動したために画角から消滅したことであると判定する．包含されている場合は次の推定を行う．その次に，歩行者 i_1 は，分裂や合体により歩行者 i_1 を派生した歩行者 i_0 が存在するかを推定する．存在する場合，歩行者 i_1 に対応する観測値が存在しない原因は歩行者 i_0 の分裂や合体が終了したために消滅したことであると判定する．歩行者 i_0 が存在しない場合，歩行者 i_1 に対応する観測値が存在しない原因は，別の歩行者 i_2 との合体が発生したために消滅したことであると判定する．

観測されたセグメントに対応する歩行者が以前のフレームに存在しない場合は，観測値 j_1 が歩行者 i の予測位置 $x_{i,k-1}$ の内部に包含されているかを推定する．包含されている場合，観測値 j_1 に対応する歩行者が存在しない原因は，歩行者 i_1 が観測値 z_{j_1} と観測値 z_{j_2} に分裂したために観測値 j_1 や j_2 が新たに出現したことであると判定する．包含されていない場合，観測値 j が画角外から画角内へ移動したために出現したこと，もしくは 2 人の歩行者 i_2 と i_3 が合体したために新しく出現したことであると判定する．

4.5 カルマンフィルタの更新

本章では，歩行者の位置予測に用いるカルマンフィルタの更新について述べる．本研究ではオクルージョンが発生した際に，その観測を利用する歩行者の位置予測に用いるカルマンゲインを小さくすることにより，オクルージョンの発生している観測値については大きい誤差を許容しつつ予測に用いることが可能となる．また分裂や合体が発生した際には，これらにより派生した歩行者および派生元となる歩行者の双方についてカルマンフィルタの更新を行う．

まず，観測値との割り当てのある歩行者のカルマンフィルタの更新について述べる．観測値との割り当てのある歩行者については，カルマンフィルタの更新時にカルマンゲインを調整することで観測値の誤差の許容範囲を変化させ，位置予測の精度を向上させている．カルマンゲインは，カルマンフィルタ更新時の，真値に対する観測値の誤差の許容される度合いを表すパラメータである．カルマンゲインを大きくすることにより，観測値と真値の誤差の許容範囲を小さく設定することができる．カルマンゲインを小さくすることにより，観測値と真値の誤差の許容範囲を大きく設定することができる．本研究では観測値が割り当てられた歩行者のカルマンゲインを決定するために，各歩行者のオクルージョンの度合いを用いる．このオクルージョン度合いを求めるため，観測値の体積の絶対値および割り当てられた歩行者の過去 5 フレームの観測値のうち最大となる体積との相対値を用いる．オクルージョン度合い O_{ocl} は以下の式で定義される．

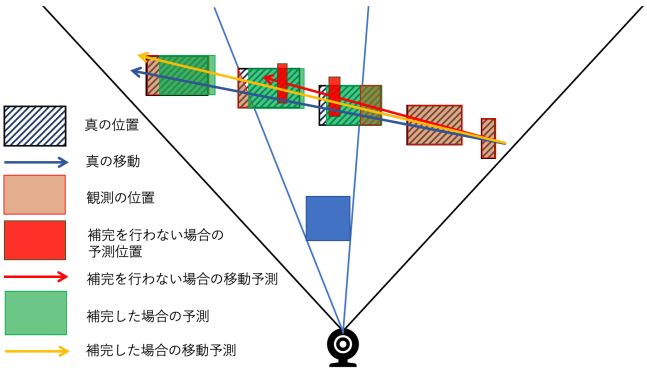


図 5 オクルージョン時に補完することでの位置予測の精度向上

$$V_{abs} = \begin{cases} V_{det}/V_{thr} & (V_{det} < V_{thr}) \\ 1 & (V_{det} >= V_{thr}) \end{cases} \quad (4)$$

$$V_{rel} = V_{det}/V_{max} \quad (5)$$

$$O_{ocl} = \beta V_{abs} + (1-\beta)V_{rel} \quad (6)$$

V_{abs} は観測値の絶対値の大きさを表しており，観測値の体積の絶対値 V_{det} が閾値 V_{thr} よりも大きい時は 1 とし，小さい時は閾値 V_{thr} に対する割合とする．V_{rel} は歩行者の相対値を表しており，その歩行者の過去 5 フレームでの体積の最大値 V_{max} に対する観測値の体積の絶対値 V_{det} の割合を表す．そして，$\beta = 0.2$ は V_{abs} と V_{rel} の利用の割合を表す．このオクルージョン度合い O_{ocl} が大きいほどカルマンゲインを小さくし，オクルージョン度合いが小さいほどカルマンゲインを大きくする．またオクルージョンをしている人については，観測値をその歩行者の x 軸方向および y 軸方向の長さの最大値になるよう補完を行い位置の予測を行う．**図 5** に，オクルージョン時に補完することでの位置予測の精度向上について示す．補完を行わずに観測値のみで位置予測を行なった場合，オクルージョン発生以降のフレームで予測位置が真の位置と大きく異なるため，同一人物であると判定することが困難となる．一方，補完を行いつつ位置予測を行った場合，予測位置が真の位置に近く判定されるため，同一人物であると判定されやすい．

次に，割り当てのない歩行者のカルマンフィルタの更新について述べる．まず，その歩行者が分裂や合体を発生さ

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

図 6 Structure Core

表 1 Structure Core の性能

項目	性能
計測可能距離	0.3 - 5m (最大 10m)
精度	±0.29%
解像度	1280 × 960
フレームレート	54 FPS
視野角	59° × 46° × 70°
消費電力	2.0W (通常時), 3.1W (最大)

せていない場合，観測値なしでカルマンフィルタを更新する．これは以前のフレームから予測される速度で移動したセグメントを観測値としてカルマンフィルタを更新するものである．次に，分裂や合体により派生した歩行者が 1 人の場合，観測なしとして更新を行う．分裂や合体により派生した歩行者が 2 人の場合，それらの歩行者の点群データの xyz 座標からこれらが合体したと仮定し xyz 座標を求め，この xyz 座標を派生後の歩行者の検出値として代用しカルマンフィルタを更新する．

5. 評価実験

提案する歩行者トラッキングに対する評価実験の結果を述べる．この評価実験では，三次元深度センサは Occipital, Inc. の Structure Core（**図 6**）を利用した．**表 1** に Structure Core の性能を示す．

評価に利用したデータは，大型商業施設のエントランスおよび大阪大学情報科学研究科研究室の 2 箇所で，実際に取得したデータを利用した．データ取得時は，センサを床から約 1m の高さに床と水平になるように設置した．利用したデータの詳細は 5.1 節で述べる．

5.1 評価用データセット

評価用に作成したデータセットの詳細を**表 2** に示す．データセットはシーン A〜F のデータからなる．それぞれについて説明する．

シーン A（**図 7**(a)）は，センサ位置からの奥行方向（センサ座標系の Z 軸正方向）において最奥部の歩行者が高さ 1m 程度の台車を押しながら左から右（同座標系で X 軸正方向）へ移動しており，その約 2m 背後に別の歩行者が追随して移動している．その手前を 3 人の歩行者が右から左へ並んで移動している．このシーンには Z 軸方向に最大 4 名の歩行者が並び，かつ最奥部の歩行者は台車を押していることからオクルージョンによりセグメント検出が極端に不安定である．シーン B（図 7(b)）も，シーン A と同様に最奥部の歩行者が高さ 1m 程度の台車を押しているが，シーン B では 5 人の歩行者が集団でその手前を通過するため，歩行者や台車のセグメント検出の不安定さはシーン A よりも高い．シーン C（図 7(c)）では，6 人の歩行者がバラバラのタイミングでセンサ観測領域に進入し，6 人のうち 2 人がスーツケースを引いて歩行した．進入方向は 3 名ずつで左および右とし，あえて進行方向が重なるように進入した．したがって各歩行者は衝突回避行動を頻繁にとった．シーン D〜F では，大型商業施設の協力を得て同施設の異なるエントランスで計測したデータである．正解データは目視により記録した．シーン D では，同時に高々 2 人の来客がセンサ観測領域を通過した．通過した歩行者には，ベビーカー利用者が 1 人，身長が 1m 程度の子供が 1 人含まれていた．これとは異なりシーン E では，身長以上の高さのある台車を運ぶ店員や，ベビーカーを押す歩行者と一緒に歩いていた 2 歳程度と推定される幼児との組など，様々な属性の歩行者がエントランスを通過した際のデータである．最後にシーン F では，画角端に駐車料金精算機があり，その利用者数人が精算機の前に列を作っていた．したがって，列が動いたり，別の歩行者が付近を通過した際のデータである．

(a) シーン A　(b) シーン B　(c) シーン C

図 7 各シーンの状況

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 [2] で導入された CLEAR 指標を採用する．CLEAR 指標は以下の評価指標を定義している．

- 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 (7) にて計算される．

$$MOTA = 1 - \frac{\sum_t (FP_t + FN_t + IDsw_t)}{\sum_t GT_t} \quad (7)$$

式 (7) 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．

- MOTP: 複数人のトラッキングにおける位置予測の精度を表す．(8) 式にて計算される．

$$MOTP = 1 - \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \quad (8)$$

式 (8) 中での d_t^i，c_t は，フレーム t の i 番目の歩行者に対応する観測バウンディングボックスと推定された

表 2 評価用データセット

名前	場所	FPS	解像度	フレーム数 (時間)	人数	検出数	密度	分裂数	合体数	一部欠損	全体欠損
シーン A	研究室	20	640 × 480	113(00:06)	5	287	2.54	2	7	84	52
シーン B	研究室	20	640 × 480	115(00:06)	6	308	2.68	2	3	126	27
シーン C	研究室	20	640 × 480	171(00:09)	6	386	2.26	3	3	64	8
シーン D	商業施設	5	640 × 480	103(00:21)	10	123	1.21	1	2	15	0
シーン E	商業施設	5	640 × 480	144(00:29)	14	285	1.17	7	2	70	8
シーン F	商業施設	5	640 × 480	109(00:21)	5	165	1.62	1	0	40	0
合計				755(01:32)	46	1554	2.06	16	17	399	95

バウンディングボックスの積（intersection）として得られる直方体の体積，およびそれらの和（union）として得られる直方体の体積をそれぞれ表す．したがって，全歩行者の全期間のバウンディングボックスのIoU（Intersection over Union）に相当する．

- 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 実験結果

それぞれのデータセットに対して精度評価を行った結果を**表 3** に示す．全データに対する MOTA は 0.914 であり，検出と割当ての精度は十分高いことがわかる．また MOTP は 0.520 であった．これは平均的に推定バンディングボックスの 3 分の 2 の体積が正解のそれに含まれているとみなすことができ，高精度に軌跡推定ができていることがわかる．この結果から，シーン A やシーン B のように複数人が団体となって移動することでオクルージョンが頻繁かつ観測領域の広範囲に発生する場合にも，オクルージョンによる欠損の補完を行うことで，オクルージョンの少ないシーンに近い精度でトラッキングできていることがわかる．また，シーン A〜D のように歩行者がその身長と比較して低い物体とともに移動している場合でも，トラッキング精度に大きな影響がないことがわかる．さらに，シーン F のように画角端で複数人がすれ違う場合も，高精度にトラッキングを行えていることがわかる．

シーン A でのトラッキング結果を**図 8** に，実際の状況を**図 9** に示す．

図 8 では，観測した三次元深度データを変換した三次元点群データを上空からの俯瞰図で表示している．青色の直線は水平画角を表しており，橙色の直線は画角から 30cm の位置を表している．また，灰色は点群を表しており，歩行者であると認識されたセグメントは，歩行者ごとに別の色のバウンディングボックスと ID で囲い表示している．この状況では，最奥部の歩行者がフレーム 1 の時点では右向きに移動する歩行者 101 として観測されている．しかし，フレーム 2，フレーム 3 で手前を左向きに移動する歩行者 103 と歩行者 104 によりオクルージョンが発生する．フレーム 4 で歩行者 103 と歩行者 106 の右側に観測された時，フレーム 2 で観測された位置からは離れた位置まで移動している．オクルージョン時も予測を行い続けることでフレーム 4 時点で歩行者 101 として再度観測できていることが確認できる．

シーン E で FP が大きくなった原因には，身長以上の高さのある台車を押す歩行者に対し，複数人であると誤判定したことがあげられる．本研究ではセグメント化を行う際に，地上から 1m 程度の高さに設置したセンサよりも高い位置の点群のみを用いている．そのために，台車の一部が歩行者として判定されたと考えられる．同様に，FN が大きくなった原因として，身長の低い子供に対応する点群がセンサよりも低かったため，セグメント化できなかったためと考えられる．

シーン F で MOTA や MOTP が高いにもかかわらず MT が低くなった原因には，複数人が接近した際に ID の入替わりが発生し，そのままトラッキングが継続されたことがあげられる．他の歩行者に接近し，その付近で立ち止まる場合，カルマンフィルタの移動予測ではこれまでの移動をもとに予測するため，その歩行者の位置に移動したとみなしがちである．本シーンもその現象が発生したものと考えられる．

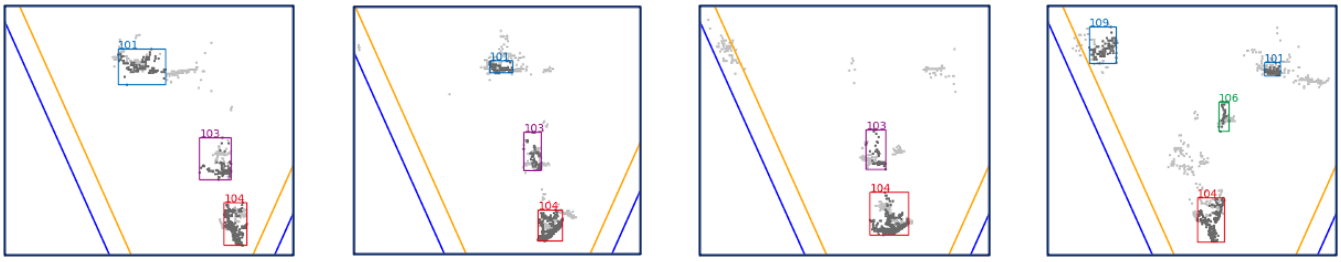
6. まとめ

本研究では，公共空間で計測した三次元深度データから，個々の歩行者をセグメント化し，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，三次元深度センサにて得られる三次元深度データを，背景差分とクラスタリングにより歩行者を表す三次元深度データのセグメントを抽出し，オクルージョンが発生している場合抽出したデータに対し補完を行った．その後，カルマンフィルタを用いてそれぞれの歩行者の位置と速度を

Vol.2021-DPS-187 No.25
Vol.2021-MBL-99 No.25
Vol.2021-ITS-85 No.25
2021/5/28

表 3 評価結果

名前	MOTA	MOTP	MT(%)	ML(%)	FP	FN	IDsw	Frag
シーン A	0.891	0.510	60.0	20.0	6	27	2	17
シーン B	0.862	0.526	66.7	16.7	14	32	2	27
シーン C	0.964	0.543	100.0	0.0	5	8	1	19
シーン D	0.921	0.506	80.0	10.0	8	1	1	2
シーン E	0.881	0.490	73.3	20.0	16	19	0	14
シーン F	0.976	0.615	60.0	0.0	1	0	3	1
合計	0.914	0.520	74.5	12.8	50	87	9	80



(a) フレーム 1　(b) フレーム 2　(c) フレーム 3　(d) フレーム 4

図 8 シーン A におけるトラッキングの一例

(a) フレーム 1　(b) フレーム 2　(c) フレーム 3　(d) フレーム 4

図 9 シーン A のトラッキングの実際の状況

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った．実際の通行データを市販の小型深度カメラを用い収集し，そのデータのうち計 755 フレームの深度データを用いた精度評価の結果，MOTA が 0.914，MOTP が 0.520 を達成した．これにより，不完全な三次元点群を用いた場合でも公共空間での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 本研究成果は国立研究開発法人情報通信研究機構（NICT）の委託研究「ウイルス等感染症対策に資する情報通信技術の研究開発（課題番号 222）」により得られたものです。

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] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.