Vol.2020-DPS-185 No.2
2020/12/21

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

右京 莉規[1]　廣森 聡仁[1]　山口 弘純[1]　東野 輝夫[1]

概要：本研究では三次元深度センサーから取得した不完全な三次元点群データを用いて，公共空間における歩行者のトラッキングを行う手法を提案する．ビル内や地下街の通路，広めのエントランスなど，複数の歩行者が多く行きかう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部分的に欠損していたり，ノイズにより実際の体格より大きく観測されるといった不完全性を有する．これに対し，提案手法では，検出された点群のクラスタサイズに基づく点群クラスタの連続性の予測とそれに基づくカルマンフィルタを適用することにより，不特定多数の歩行者が観測領域に進入・退出を行い，相互にセンサー視野を頻繁に遮蔽する環境において，三次元点群の時系列観測に対する堅牢な軌跡特定を実現する．実商業施設のエントランスにおいて市販の小型深度カメラを用いて収集した，様々な属性の歩行者を含む深度データを用いた実証実験において，軌跡同定に関する適合率 92.7%，再現率 94.6%を実現した．

1. はじめに

公共施設において，施設の来訪者数を正しく知ることは重要である．例えば，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，滞在者数を常時把握しておくことで，効率的な避難誘導や安全確認も可能となる．また来訪者数は，商業施設等のマーケティングのみならず，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きわめて重要である．

こういった来訪者把握の需要に対し，近年では RGB 画像から人物を検出する方法があるものの，防犯用途以外では制約も多く，オプトアウトの申立てに対する対応作業の負荷も課題となることから，特に公共空間において RGB 画像による人物検出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ーを用いて，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．三次元距離センサー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，センサー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．

しかし，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．本来，人流検知を目的としたセンサーは複数の人物の存在においてもオクルージョンが発生しにくい箇所に設置することが望ましい．にも関わらず，実空間におけるセンサー設置は電源の確保や安全性の保証，美観を損失しないことや看板を遮蔽しないことなどから制約を受けることがほとんどであり，既存の建造物に対する設置個所の自由度は極めて低い．したがって，オクルージョンが頻繁に発生するようなセンサー視野角においても人流を確実に追跡できる技術が求められる．

本研究では三次元深度センサーから取得した不完全な三次元点群データを用いて，公共空間における歩行者のトラッキング（軌跡導出）を行う手法を提案する．提案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，歩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歩行者に対応するセグメント抽出を行う．各セグメントの連続性から移動推定を行セグメント化を行った後，カルマンフィルタを適用することにより歩行者の軌跡導出を行う．

提案手法は，ビル内や地下街の通路，広めのエントランスなど，複数の歩行者が多く行きかう空間を対象とし，三次元深度センサーは通路脇などの壁面に設置される一般的な状況を想定する．こういった環境においては，センサー

[1] 大阪大学大学院情報科学研究科

座標系の前方（奥行き）方向が歩行者の主な移動方向と直交するため，近傍を通過する歩行者がより遠方を通過する歩行者を遮るオクルージョンが多く発生する．これに対し，提案手法では，検出された点群のクラスタサイズに基づく点群クラスタの連続性の予測とそれに基づくカルマンフィルタを適用することにより，不特定多数の歩行者が観測領域に進入・退出を行い，相互にセンサ視野を頻繁に遮蔽する環境において，三次元点群の時系列観測に対する堅牢な軌跡特定を実現する．

実際の商業施設のエントランスにおいて市販の小型深度カメラを用い，約 7 時間にわたり収集した深度データを用いた精度検証の結果，歩行者検出において適合率 99.5%，再現率 89.5%，F 値 94.3%を達成し，軌跡同定に関しては適合率 92.7%，再現率 94.6%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．

本章の構成は以下のとおりである．2 章では関連研究について紹介し，本研究の位置付けについて明確にする．3 章では，提案手法の概要について述べ，4 章では，提案手法の詳細について説明する．5 章では，実際の商業施設で実施した実験とその評価について述べる．そして，6 章で本研究のまとめについて述べる．

2. 関連研究

スマートフォンに代表されるモバイル端末に搭載された Wi-Fi や Bluetooth，慣性センサー等を用いた端末位置推定技術は現在においても多数の研究がみられる．最近のものでは，例えば文献 [1] では，無線が不安定な環境においても，高精度で位置推定が可能な手法として，WiDeep が提案されている．WiDeep は，受信した Wi-Fi 信号を，オートエンコーダのモデルと確率的フレームワークにより処理することで，スマートフォンとその場所の Wi-Fi 信号の関係を導出し，高精度の位置推定を実現している．一方，端末を保持しない不特定多数の歩行者を LiDAR などの設置型センサーを用いて追跡する手法も盛んにおこなわれている．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，現在，大型ショッピングモールに本システムを展開している [2].

近年，三次元深度センサーの普及に伴い，センサーにより得られる三次元点群データを対象に，物体の種類や人の骨格など，対象物の存在だけでなく，そのコンテキストを推定する取組みが注目を集めている．文献 [4] では，ディープニューラルネットワークにより，入力として与えられた三次元点群データに対し，その物体が何であるかを判定するクラス分類と，ある物体を個々の部位に分割するセグメンテーション分類を実現する PointNet が提案されている．文献 [5] では，より現実的な環境において，物体を認識するために，異なる粒度の三次元点群データを扱えるよう，前述の PointNet を階層的に適応することにより，様々な物体が配置された複雑な状況においても，物体を認識できる手法が提案されている．また，VoteNet [6] では，三次元点群データに対し，ハフ変換を活用した投票により，物体の中心位置を予測し，これに基づき，物体に対するバウンダリボックスを高い精度で導出する手法を提案している．同様に，文献 [7] においても，三次元点群データ中の物体に対するバウンディングボックスを予測することにより，個々の物体に紐づく三次元点群データを抽出する手法が提案されている．

一方，物体認識のみならず，人の姿勢を把握する手法も近年盛んに研究されている [8], [9], [10], [11]．Complex YOLO [10] は，取得した三次元点群データ群を俯瞰視点に基づく二次元画像に変換し，物体検出の手法を適応することで，物体の存在を把握する手法である．文献 [9] においては，三次元点群データを 3D ボクセルに変換することにより，個々の形状特徴を維持しながら手や体の姿勢を推定する手法を提案しており，人物を前面と上面から撮影したいずれの場合でも高精度に全身の姿勢を推定できることを示している．文献 [12] では，入力となる三次元点群データに対し，予め定義された姿勢のうち，尤もらしい姿勢を高速に導出する手法を提案している．また，慣性センサーと単一視点深度カメラからの情報を融合することにより，高速な動きに追随し，リアルアイムに姿勢を推定する手法も提案されている [11]．三次元点群を前提としたこれらの手法は，対象物の存在や人物の姿勢検出を主目的としており，提案手法のように，複数の歩行者が存在し移動するもとでの軌跡推定を行うものではない．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 [13]．一方で，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 MOT の試みは現状でほとんどみられていない．

3. システムアーキテクチャと動作概要

本研究は，三次元深度センサーから記録される三次元深度データを解析することにより，様々な属性の移動制約者を含む公共空間の歩行者のトラッキングを行う．本研究のフローチャートを**図 1** に示す．本システムの入力には，三次元深度センサーから得られた三次元深度データを用いる．その入力データを実際の座標系に変換し，あらかじめ取得しておいた背景についての三次元深度データを用いて背景差分を行うことで移動物体の三次元点群データのみを抽出する．その後，抽出した三次元点群データを俯瞰視点の二次元点群データに変換しクラスタリングを行い，ノイズ除去と歩行者のセグメント化を行う．公共施設の場合複数の歩行者が接近することがあるが，その際に接近する部位は

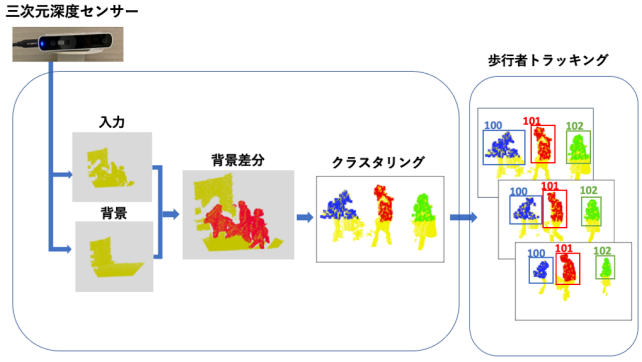


図 1 歩行者トラッキングフローチャート

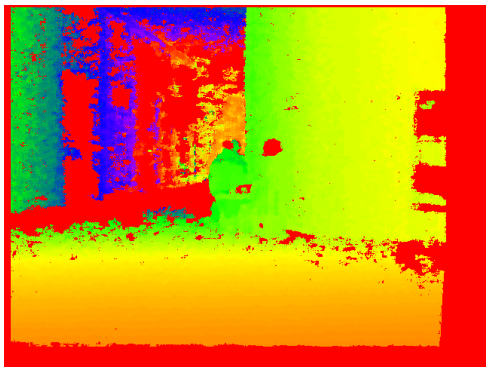
手や足などの末端が多い．複数人が接近した場合，クラスタリングで 1 人であると判定されることがある．このような場面でも 1 人ずつのセグメントを判定するために，本システムでは比較的接近しにくい上半身の点群データを用いてクラスタリングを行う．そして，それぞれのセグメントに対し，それ以前のフレームで捕捉された歩行者のセグメントと同一人物を表すかどうかを判定し，カルマンフィルタを用いて歩行者の次フレームでの位置を予測する．そのデータに含まれる歩行者についてトラッキングを行い，同一人物であるかを判定する．

4. 人流検知と歩行者のトラッキング

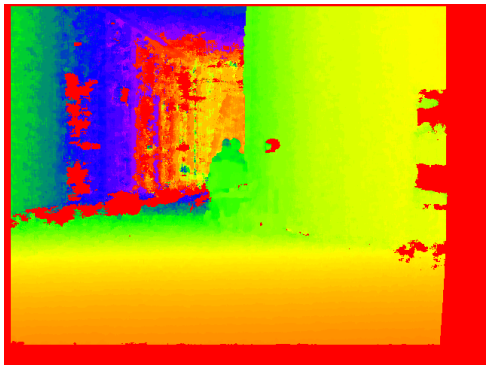
本章では三次元深度センサーにて捕捉された歩行者のトラッキング手法について述べる．まず，センサーにて取得したデータを，4.1 章で述べる背景差分を行うことで移動物体を抽出する．次に，4.2 章で述べるクラスタリングを行うことで，それぞれの歩行者のセグメント化を行う．取得したデータに歩行者が含まれない場合，そのデータを棄却する．その後，それぞれの歩行者のセグメントの位置をカルマンフィルタを用いて予測する．現在のフレームで検出されたセグメントとそのフレーム以前に検出された歩行者のセグメントの対応を線形割り当てを行うことで求め，次のフレームでのその歩行者の位置を予測することによりトラッキングする方法を述べる．

4.1 背景差分による移動物体の抽出

本システムでは，背景差分を行うことで移動物体を抽出している．本システムで行う背景差分は以下の通りである．まず背景データとして，三次元深度センサーで捕捉可能な領域に対し，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．深度センサーが取得する三次元深度データには，デバイス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 30 フレームの三次元深度データを取得し，各ピクセルの中央値を求めることでこの欠損値を除去する．**図 2(a)** は 1 フレームから生成した背景データとなっており，**図 2(b)** は



(a) 1 フレームを利用



(b) 30 フレームの中央値を利用

図 2 利用するフレーム数による背景データの誤差の違い

30 フレームの中央値から生成した背景データを表している．なお，赤色で示されているピクセルが欠損値を表している．これらの図から，30 フレームの中央値を取ることで欠損の影響を抑えられていることが確認できる．

移動物体がセンサー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が背景となる三次元深度データとの間に差分が生じる．その差分が生じた領域のみのデータを抽出することにより，移動物体の三次元深度データを取得することができる．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，本研究では取得したデータと背景となるデータの差分が 10cm 以上となる領域のみを移動物体として判定する．

4.2 クラスタリングによる移動体セグメントの抽出

背景差分により抽出した点群データを自動クラスタリングすることにより，異なる複数の移動物体を分割し，またデバイスの測定誤差により発生するノイズを除去する．自動クラスタリングではまず，背景差分により抽出された点群データに対し，データ取得に利用する三次元深度センサーの最大検出距離以上のデータを削除する．本研究で用いた三次元深度センサーの最大検出距離は 5 メートルのため，本システムでは 5 メートルよりも遠くのデータを削除する．次に，処理時間を短縮するため，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．

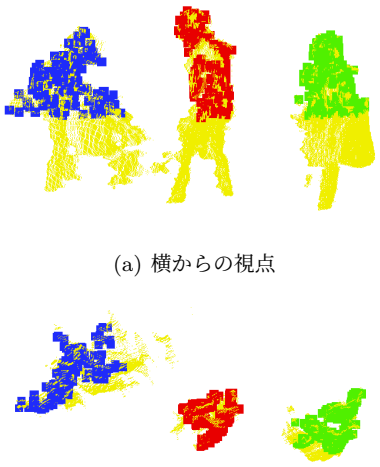
その後クラスタリングを行い，歩行者の点群データをそれぞれのセグメントに分割する．クラスタリング時の処理時間を短縮するために，鉛直方向の軸を取り除き三次元点群を二次元点群に圧縮する．三次元点群に対しクラスタリングを適用する手法も考えられるが，公共施設内に存在する移動物体は二次元平面を移動する人およびその人が所持，利用しているものであるため，鉛直方向に複数人存在することを考慮する必要がない．そのため，処理時間短縮のために鉛直方向の情報を削除することが可能となる．

クラスタリング対象となる点群が存在する場合，得られた二次元点群に対し DBSCAN アルゴリズムを用いたクラスタリングを行うことで，ノイズを除去し対象となる移動物体を分割する．DBSCAN アルゴリズムは，点群の密度に基づきクラスタリングを行う．本研究の対象となる移動物体は，形状や大きさ，数が不定である．DBSCAN アルゴリズムはこのようなデータに対し堅牢であるため，移動物体のデータを容易に抽出することが可能となる．また，本研究ではクラスタリングに用いる点群データを，データ取得に用いた三次元深度センサーよりも高い位置に存在する点群のみとしている．公共施設では，複数人がセンサーの検出可能領域で接近することがあるが，このような場合に歩行者ごとのセグメント分割の精度が低下することがある．接近する部位は手や足などの末端が多い．公共空間では歩行者はほとんどの場合手を下ろしているため，接近しやすい末端箇所は下半身の高さに多い．そのため，本研究では比較的接近しにくい上半身を表す点群データを用いてクラスタリングを行う．**図 3** はクラスタリングを適用しセグメントごとに分割した点群データを表している．

4.3 カルマンフィルタによるトラッキング

クラスタリングにより抽出したセグメントが，それまでに出現した歩行者と同一人物かを線形割り当てを用いて判定し，カルマンフィルタを用いて歩行者ごとに次フレームでの位置を予測する．本研究で用いるカルマンフィルタでは，歩行者は位置を表す xyz 座標および xyz 方向の速度の 6 つの情報を持つ．セグメントの位置を表す xyz 座標は，それぞれの方向に対する最大値と最小値の平均をとったものを中央の座標とし用いる．ここでの xyz 軸の定義は**図 4** にて示す．本研究での同一人物判定アルゴリズムについて，**図 5** に示す．

図 5 に示すように本システムでのトラッキングでは，まずそのフレームの直前に歩行者が存在していたかどうかを調べる．つまり，その時点でトラッキングを行なっている歩行者がいるかどうかを確認する．存在しない場合，検出したセグメントを新しい歩行者として，次フレーム以降で



(a) 横からの視点

(b) 上からの視点

図 3 クラスタリングにより得られた点群のセグメント．背景差分により抽出された点群データを黄色で表し，クラスタリングにより得られた点群データを青色，赤色，緑色で表している．

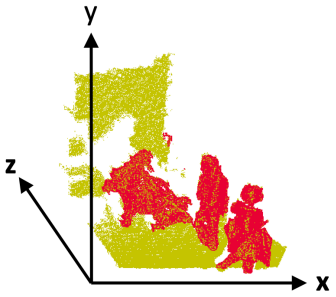


図 4 xyz 軸の定義

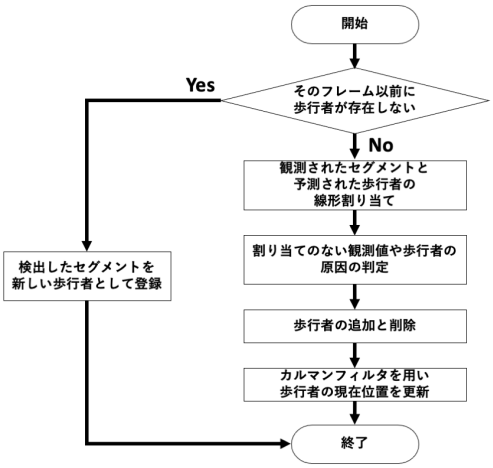


図 5 同一人物判定アルゴリズム

の位置予測を可能にする．この際，次フレームでの位置はこのフレームで検出した位置とし，速度は xyz 方向の全てを 0 として位置の予測を次フレーム以降で行う．

直前のフレームに歩行者が存在している場合，以下の処理を行う．まず，そのフレームで観測された歩行者のセグ

メントと，直前に存在している歩行者の対応を線形割り当てにより調べる．観測された全てのセグメントおよび前フレームで予測された全ての歩行者の組に対し，位置および体積の差を求め，その差から求めたそれぞれの組に対応するコストの和が最小になるように 1 対 1 に割り当てる．線形割り当てに用いるコスト c は以下の式の通りに求める．

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

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

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

この式での $\alpha = 0.8$ 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 i と観測値 j の距離を表す．また，$V_{i,j}$ は歩行者 i と観測値 j の体積の差を表す．x_j, y_j, z_j, V_j はそれぞれ観測されたセグメントの xyz 座標および体積を表し，x_i, y_i, z_i, V_i はそれぞれ予測された歩行者の xyz 座標および体積を表す．線形割り当てでは，このコストの和を最小にする．なお，観測されたセグメントの数と予測された歩行者の数が異なる場合は，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．

予測された歩行者や観測されたセグメントのうちに割り当ての存在しないものが存在する場合，その原因について推定を行う．原因は 4 種類存在し，歩行者がセンサーの画角外から画角内に移動したために観測されたこと，歩行者が画角内から画角外へ移動したために観測されなくなったこと，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て判定されたこと，および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 つに合体して判定されたことである．これらの判定方法については 4.3.1 章にて詳しく述べる．この判定により新しく画角内に出現した歩行者や，分裂の発生により新しく生成されたセグメントについては，次フレーム以降での位置予測を行う．また，2 フレーム以上連続で観測値との割り当てのない歩行者については画角外へ消滅したと判定し，次フレーム以降での位置予測を行わない．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位置予測を行う．分裂や合体が終了し元の歩行者に対応する観測値を 2 フレーム以上連続で取得すると，分裂や合体により発生したセグメントの位置予測を終了する．なお分裂や合体の発生時は，分裂や合体元の歩行者に対応する観測値が 2 フレーム以上連続で存在しない場合でも位置予測を続ける．その際，分裂や合体後の歩行者の観測値を用いて元の歩行者のカルマンフィルタも更新する．

その後，割り当ての存在する歩行者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．この際，割り当てられている観測値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 4.3.2 章で述べる．このように現在の位置を更新し，現在のフレームについての処理を終える．

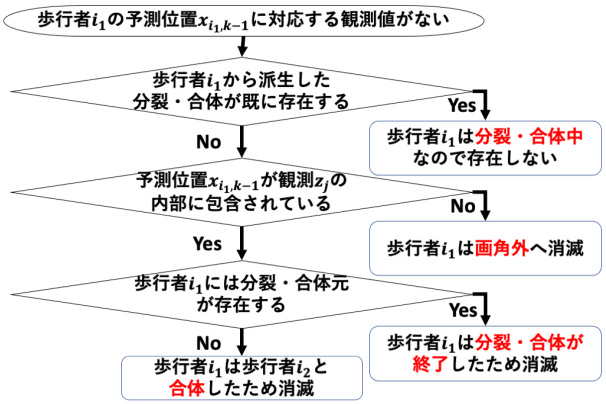


図 6 歩行者 i_1 に対応する観測値が存在しない場合の原因推定アルゴリズム

4.3.1 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値が存在しない場合と観測値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値が存在しない場合について述べる．この場合の原因推定アルゴリズムを**図 6** にて示す．まず，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．

次に，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について述べる．この場合の原因推定アルゴリズムを**図 7** にて示す．この場合は，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

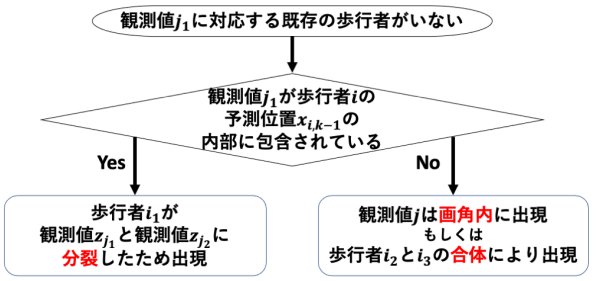


図 7 観測値 j_1 に対応する既存の歩行者が存在しない場合の原因推定アルゴリズム

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．

4.3.2 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．本研究では複数人が同時にセンサーのデータ取得可能範囲を通過することがある．そのため，1 人の歩行者がセンサーのデータ取得範囲を通行することで死角となる範囲を別の歩行者が通行することにより，オクルージョンが発生することがある．その際に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．

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

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

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

図 8 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)

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．

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．なお観測値の場合でのカルマンフィルタの更新は，以前のフレームから予測した速度での移動が行われたとして行う．次に，分裂や合体により派生した歩行者が 1 人の場合，派生した歩行者から発生する分裂や合体が存在するかを推定する．存在する場合，派生元の歩行者のカルマンフィルタを観測値なしで更新する．存在しない場合，合体の発生直後のフレームの場合のみは観測値なしで更新を行い，そうではないフレームでは派生後の歩行者の位置の変化を派生元の歩行者に同様に適用し，カルマンフィルタを更新する．また，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た際の 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

本章では，提案する歩行者トラッキングに対する評価実験とその結果について述べる．この評価実験では，三次元深度センサーは Occipital, Inc. の Structure Core（**図 8**）を利用した．**表 1** に Structure Core の性能を示す．大型商業施設のエントランスに，床から約 1m の高さで床と水平になるように，このセンサーを設置し，施設の来客がエントランスと店内の間を通過する様子を約 7 時間計測した．この計測では，毎秒 5 フレームで点群データを計測し

表 2　個々のフレームに対する歩行者検出結果

項目	精度 (%)
適合率	99.5
再現率	89.5
F 値	94.3

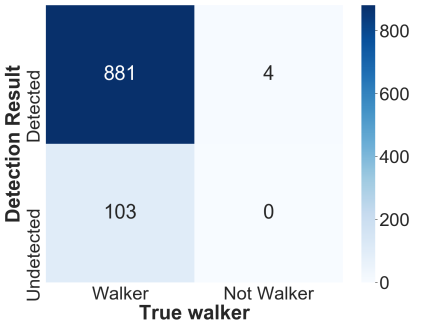


図 9　個々のフレームに対する歩行者検出結果の混同行列

ており，収集した点群データのうち，約 3 分間の計測に相当する，計 851 フレームの点群データに対し，人の手でラベル付けを実施した結果，1 つのフレームあたりに含まれる歩行者数は 0 人から 5 人で，歩行者の合計数は 32 人であった．このラベル付けした点群データに対し，このシステムを用いてトラッキングによる推定精度を評価する．

5.1　個々のフレームに対する歩行者検出率

まず，個々のフレームに対する歩行者検出の精度を評価する．ここでは，851 フレームの点群データに対し，個々の歩行者をセグメントとして抽出できているか否かを評価した．なお，2 人以上の歩行者が 1 つのセグメントとして検出された場合は，いずれの歩行者も検出できていると判定する．個々のフレームに対する歩行者検出の適合率，再現率，F 値を**表 2** に，それらの内訳を**図 9** に混同行列として示す． 表 2 に示されるように，適合率は高く，誤って歩行者を検出していないことがわかる．一方で，再現率は 9 割に満たず，提案手法により検出できていない歩行者が一定数存在していることがわかる．提案手法では，三次元深度センサーよりも高い位置に存在する点群のみを対象としているため，身長の低い子どもを検出できず，また，荷物を保持していた歩行者が，途中でその荷物を床に置くことで手放した際，同一の人の継続して把握できないことが挙げられる．

5.2　歩行者に対するトラッキング精度

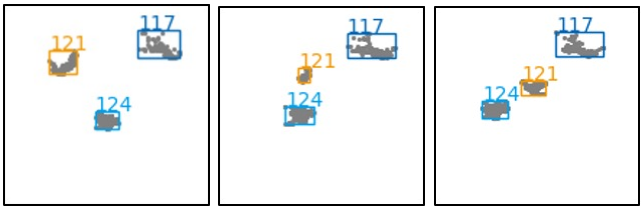
同様に，851 フレームの点群データに対し，二つの観点から，歩行者に対するトラッキング精度を評価する．一つ目の観点は，軌跡同定に関する適合率で，提案手法により，同一の人であると判定されたデータが，別の人のデータを含んでいないことを評価するためのものである．提案手法は個々の歩行者を区別できるよう異なる ID を割り当てるが，歩行者の判定が正しく為されずに，一つの ID に複数の歩行者に割り当てられる場合には，その ID に割り当てられている歩行者のうち，フレーム数が最多となる歩行者がその ID を代表する歩行者とする．ある ID が出現しているフレーム数に対し，その ID を代表する歩行者に割り当てられているフレーム数を，その ID に対する同定精度とし，これらを平均したものを軌跡同定に関する適合率とする．二つ目の観点は，軌跡同定に関する再現率で，ある歩行者に対して，その歩行者が検出されていないフレームがないことを評価するためのものである．提案手法により，歩行者の判定が正しく為されず，ある歩行者に対し，複数の ID が割り当てられている場合には，その歩行者に割り当てられた ID のうち，フレーム数が最多となる ID をその歩行者を代表する ID とする．ある歩行者が出現しているフレーム数に対し，その歩行者を代表する ID がその歩行者に割り当てられているフレーム数を，その歩行者に対する同定精度とし，これらを平均したものを軌跡同定に関する再現率とする．個々の ID に紐づくフレームにおいて観測される最大人数毎での適合率を**表 3** に，個々の歩行者に紐づくフレームにおいて観測される最大人数毎での再現率を**表 4** に示す． これらの評価結果からわかるように，適合率と再現率は高く，歩行者として検出できたセグメントに対し，カルマンフィルタにより，正確に位置を予測できていることがわかる．しかしながら，表 3 に示されるように，フレーム内に存在する人数が増加するほど，適合率が低下している傾向が確認され，特に，フレーム内の最大人数が 4 人の場合，適合率は 9 割を大きく下回っている．この原因として，センサーの計測可能範囲に，複数人が同時に出入りする際に，範囲内へ移動した歩行者を，範囲外へ移動した歩行者と誤判定してトラッキングし続けたこと，また複数人がすれ違う際に ID が逆になっていることが挙げられる．

表 3　軌跡同定に関する適合率

最大人数	該当する ID の数	平均精度 (%)
1	13	100
2	20	93.8
3	23	89.8
4	11	79.7
合計	66	92.7

表 4　軌跡同定に関する再現率

最大人数	該当する歩行者の数	平均精度 (%)
1	6	100
2	11	97.0
3	14	91.7
4	9	92.6
合計	40	94.6



(a) (b) (c)

図 10 歩行者トラッキングの一例

また，複数のフレームに対し，トラッキング処理を適応した結果を，俯瞰図にて示したものを**図 10** に示す．この図に記載されている数字は歩行者の ID を表しており，ID とともに枠線で囲まれた灰色の箇所が，抽出された歩行者のセグメントを表している．これらの図には，3 人の歩行者が存在しており，図の下の歩行者から順に，左方向へ，右方向へ，左方向へ移動している．中央に位置する歩行者が，(a) では下の歩行者よりも左側に位置していたが，(b) で下の歩行者の奥を通過し，(c) で右側へ移動している．これらの図から，異なる 3 人の歩行者に対し，異なるトラッキング ID を割り当てており，それぞれの歩行者を適切に追跡できていることがわかる．

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，個々の歩行者に対し，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，三次元深度センサー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する．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う．実際の商業施設のエントランスにおいて市販の小型深度カメラを用い，施設の来客の通行データを約 7 時間にわたり収集した．深度データを用いた精度評価の結果，歩行者検出において適合率 99.5%，再現率 89.5%，および　F 値 94.3%を達成し，軌跡同定に関しては適合率 92.7%，再現率 94.6%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

本研究は東北大学電気通信研究所共同プロジェクト研究により実施されたものである．

参考文献

[1] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef. Wideep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning. In *Proceedings of 2019 IEEE International Conference on Pervasive Computing and Communications*, pp. 1–10, 2019.

[2] 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.

[3] M. Alzantot and M. Youssef. UPTIME: Ubiquitous Pedestrian Tracking Using Mobile Phones. In *Proceedings of 2012 IEEE Wireless Communications and Networking Conference*, pp. 3204–3209, 2012.

[4] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85, 2017.

[5] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pp. 5099–5108. Curran Associates, Inc., 2017.

[6] C. R. Qi, O. Litany, K. He, and L. Guibas. Deep Hough Voting for 3D Object Detection in Point Clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9276–9285, 2019.

[7] Bo Yang, Jianan Wang, Ronald Clark, Qingyong Hu, Sen Wang, Andrew Markham, and Niki Trigoni. Learning Object Bounding Boxes for 3D Instance Segmentation on Point Clouds. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pp. 6740–6749. Curran Associates, Inc., 2019.

[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time Human Pose Recognition in Parts from Single Depth Images. In *Computer Vision and Pattern Recognition 2011*, pp. 1297–1304, 2011.

[9] Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] Martin Simony, Stefan Milzy, Karl Amendey, and Horst-Michael Gross. Complex-YOLO: An Euler-Region-Proposal for Real-time 3D Object Detection on Point Clouds. In *The European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[11] Z. Zerong, Y. Tao, L. Hao, G. Kaiwen, D. Qionghai, F. Lu, and L. Yebin. HybridFusion: Real-Time Performance Capture Using a Single Depth Sensor and Sparse IMUs. In *Proceedings of The European Conference on Computer Vision*, 2018.

[12] Manuel Marín-Jiménez, Francisco Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 3D Human Pose Estimation from Depth Maps Using a Deep Combination of Poses. *Journal of Visual Communication and Image Representation*, Vol. 55, , 07 2018.

[13] Multiple Object Tracking Benchimark. `https://motchallenge.net/` (参照 2020-11-26).