

多層 3D LiDAR と LSTM を用いた距離・点群密度変化に頑健な歩容認証

○ Jeongho Ahn 山田 弘幸 中嶋 一斗 倉爪 亮 (九州大学)

概要

本稿では、これまでに提案した 3D LiDAR と LSTM を用いた歩容認証 [7] に対し、対象の移動方向やセンサとの距離及び点群密度の変化に頑健な新たな手法を提案する。提案手法では、3D LiDAR から取られた時系列の点群を用いて歩行者の瞬間的な移動方向ベクトルを計算し、歩行者の側面方向からの計測点群に変換することで、移動方向の変化に対応する。また、3D LiDAR は対象までの距離が長くなると、取得される点群密度が低下することから、距離による点群密度の低下を考慮した新たな LSTM ベースのネットワークを構築する。これにより、センサとの距離及び点群密度の変化に対する頑健性の向上を図る。実験では、様々な移動方向とセンサからの距離に対する歩行者の点群データを含むデータセットを構築し、提案手法の有効性を検証する。

1. 初めに

歩容認証は、指紋や顔などと同様な生体認証の一つであり、遠隔から個人の識別が可能である利点を有する。特に、対象者に特別な動作を行わせる必要なく非接触で識別可能なことから、監視や入室管理などに有望な個人認証技術として期待されている。

歩容認証には、主に歩行画像を撮影するカメラが用いられることが多いが、深度情報を活用した手法 [1] も提案されている。深度情報を扱う代表的なセンサである RGB-D カメラは、照明条件や衣服の変化に対しロバストであるなどのメリットがある。しかし、測定できる距離に制限があり、屋外の計測には不向きである。

一方、深度情報が得られるセンサの一つである 3D LiDAR は、比較的遠い距離の計測が可能であることから、屋外環境での活用が期待できる。ただし、現在の多くの 3D LiDAR は水平方向に比べ垂直方向の空間分解能が極めて低い特徴があり、歩容認証に活用される例は少ない [5][6]。

先行研究 [5] では、歩行者画像のシルエットの平均である Gait Energy Image (GEI) [4] を用いた深層学習モデルが提案されている。しかしこの手法では、歩行の時系列情報は GEI に圧縮され、陽に用いられていない。また、先行研究 [6] では、歩行者の点群密度を向上するリサンプリングの手法が提案されている。これに対し、我々は時系列の点群情報に対して LSTM を用いた識別ネットワークを構築し、センサからの距離が一定かつ同一方向を移動する歩行者に対して、3D LiDAR を用いた人物識別手法 [7] を提案してきた。

本論文では、歩行者の移動方向と点群密度が変化する状況において、これまでの提案手法 [7] による識別性能の評価と、点群密度の変化に対するロバスト性の向上を目指した新たなネットワークを提案し、性能を比較した

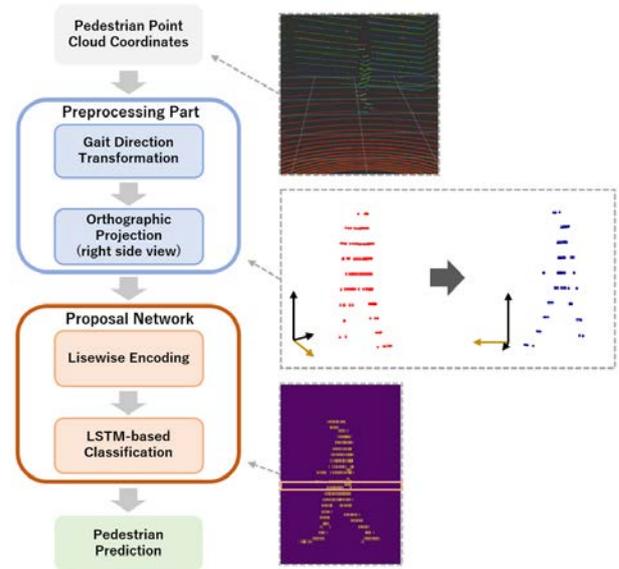


図 1 提案手法の概要

結果を報告する。

図 1 に提案手法の概要を示す。3D LiDAR から取られた歩行者の点群データから、まず前処理として歩行者の移動方向を推定し、左側面から見える 2 値画像データに変換する。2 値画像を、後述する識別ネットワークに入力し、歩行者を推定する。

2. 点群データセット構築

2.1 点群データの取得

これまでの研究 [7] では、点群情報による歩容認証の基礎的性能を評価するために、センサから見える歩行者の移動方向やセンサとの距離（点群密度）が変化しない状況でデータセットを構築し、評価した。これに対し本報告では、移動方向や点群密度の変化を含むデータセットの追加構築を行った。リアルタイムに計測可能なマルチライン 3D LiDAR (HDL-32E, Velodyne 社製) を用い、図 2 に示す計測環境において 31 人の歩行者の距離データを取得した。具体的には、被験者にセンサから 8.5m 離れた地点を中心として半径 5m の円周上を歩行させることで、センサから観測する歩行者の距離データに 360° の移動方向の変化と 3.5m から 13.5m までの距離変化が含まれる点群情報を取得した。

2.2 点群密度の異なるデータセット構築

点群密度の異なるデータに対する識別性能を確認するため、取得した点群データを点群密度及び移動方向が異なるように分割し、4 つのデータセットを構築した。具体的には、以下の方法で点群密度の異なるデータに分

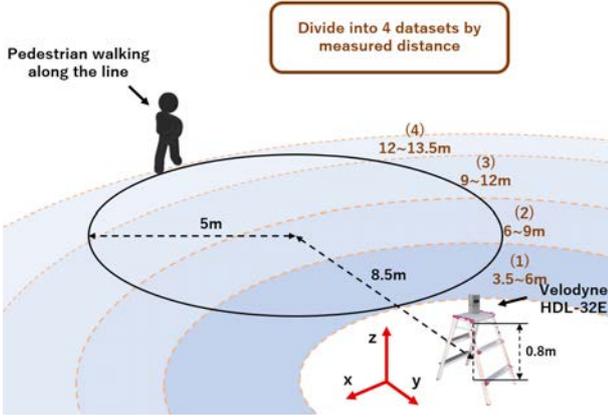


図2 歩行者の距離データ取得環境とデータセット分割

割する.

1. 時刻 t において, 背景差分によって歩行者以外の背景点群を消去し, 対象者を示す点群 $\mathbf{P}_t = \{\mathbf{p}_{t,1}, \mathbf{p}_{t,2}, \dots, \mathbf{p}_{t,N}\}$, ($\mathbf{p}_{t,n} \in \mathbb{R}^3$, N =歩行者点群の総数) を抽出する
2. 抽出された歩行者点群の中心位置 $\mathbf{c}_t = (c_{t,x}, c_{t,y}, c_{t,z})$ を以下の通り求める

$$\mathbf{c}_t = \frac{1}{N} \sum_{n=1}^N \mathbf{p}_{t,n}, \quad (\text{ただし}, c_{t,z} = 0) \quad (1)$$

3. 時刻 t における歩行者位置 \mathbf{c}_t とセンサの設置位置との距離 d_t を求める
4. 距離 d_t が 3.5m 以上 6m 未満である点群をデータセット 1 に追加し, 同様に 6m~9m の点群をデータセット 2, 9m~12m をデータセット 3, 12~13.5m をデータセット 4 に追加する
5. 1~4 を全ての時刻, 及び全ての被験者に対して行う

データセットの分割に関しては, 実際に取得したデータを観察し, 6, 9, 12m 付近で垂直方向のライン数に変化が見られたため, これらをデータセット分割の基準値とした. データセット 1~4 はそれぞれ異なる点群密度によって構成され, データセット 1 が最も点群密度が高く, データセット 4 が最も点群密度が低い. これは, センサから水平垂直方向に放射状にレーザーが照射されるため, センサからの距離が長くなると, 水平方向の同一ライン上の計測点の間隔, および垂直方向のライン間隔がそれぞれ広がるためである. 例えば, 図 3 に示すように, 最もセンサに近いデータセット 1 ではほぼすべてのラインに歩行者のデータが含まれるが, データセット 4 では歩行者は数ラインで計測されるのみである. また, データセット 1 と 4 はほぼ歩行者の側面が計測されているが, データセット 2 と 3 は歩行者を斜め前方あるいは斜め後方から計測したデータで構成されており, それぞれのデータセットで移動方向が異なる.

3. 歩行方向の変化への対応

構築したデータセットには歩行方向の異なる点群データが含まれる. そこで, これまでに提案した手法 [7] と同様に, 識別ネットワークには常に歩行者に対して同じ方向 (歩行者の左側面) が見える形に以下の方法で変換する.

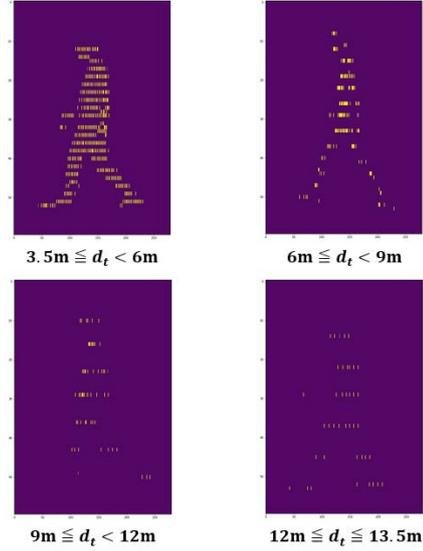


図3 センサとの距離 d_t に対する点群密度の変化

1. 時刻 t における歩行者の移動方向 θ_t を, 時系列の前後の歩行者位置 \mathbf{c}_{t-1} と \mathbf{c}_{t+1} から求める

$$\theta_t = \arctan \frac{c_{t+1,y} - c_{t-1,y}}{c_{t+1,x} - c_{t-1,x}} \quad (2)$$

2. 求めた歩行者の移動方向 θ_t の分だけ, 歩行者点群を \mathbf{c}_t を中心とし z 軸周りに回転させた新たな歩行者点群 $\hat{\mathbf{P}}_t = \{\hat{\mathbf{p}}_{t,1}, \hat{\mathbf{p}}_{t,2}, \dots, \hat{\mathbf{p}}_{t,N}\}$ を得る

$$\hat{\mathbf{p}}_{t,n} = \mathbf{R}_z(-\theta_t) \cdot (\mathbf{p}_{t,n} - \mathbf{c}_t), \quad (n = 1, \dots, N) \quad (3)$$

ただし, \mathbf{R}_z は以下の通り z 軸周りの回転行列を表す.

$$\mathbf{R}_z(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

3. 点群 $\hat{\mathbf{P}}_t$ を, 図 3 に示すような, 高さ 2.4m の範囲を 40 分割, 幅 1.4m の範囲を 140 分割した 2 値画像 $\mathbf{I}_t \in \mathbb{R}^{40 \times 140}$ に変換する

4. ネットワーク構造

これまでの研究 [7] で提案したネットワークは, 点群密度の変化を積極的に考慮したものではなかった. そこで本報告では, まずこれまでに提案したネットワーク [7] が点群密度の変化に対してどの程度の識別性能を有するかを確認した. 次に, 点群密度の変化にロバストであると考えられる新たなネットワークを構築し, その性能を検証した.

4.1 これまでのネットワーク [7] (ネットワーク A)

図 4 にこれまでに提案したネットワーク構造を示す. 提案ネットワークは CNN ベースのエンコードパートと LSTM ベースの分類パートから成る. エンコード

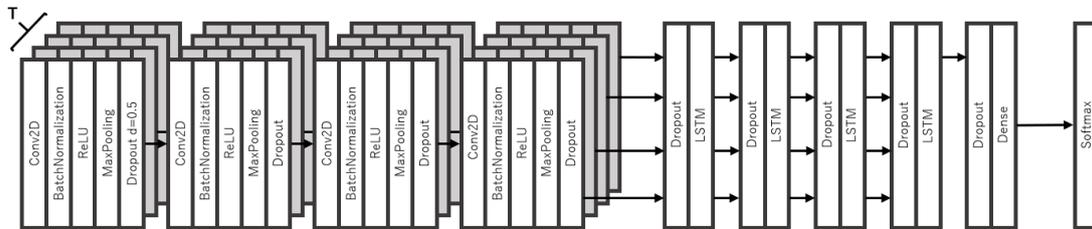


図4 これまでに提案したネットワーク構成 [7] (ネットワーク A)

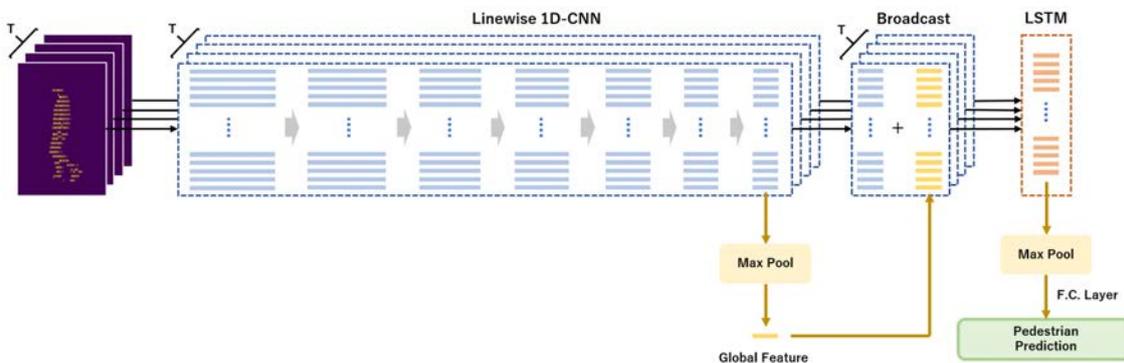


図5 新たに提案したネットワーク構成 (ネットワーク B)

パートは4層の2D-CNNと2D-Max Poolingで構成され、瞬間的な全体の歩容形状を特徴ベクトルにエンコードする。分類パートは4層のLSTMで構成され、エンコードされた T 回の観測分の特徴ベクトルにより、時系列の特徴量の変化から人物を識別する。過学習を防止するため、各CNNやLSTM間にDropout層を設けている。

4.2 点群密度の変化に頑健なネットワーク (ネットワーク B)

図5に今回新たに提案するネットワーク構造を示す。提案ネットワークはPointNet[8]をベースにラインワイズに拡張した構成である。入力される2値画像の高さ方向のライン毎に1D-CNNでエンコードし、各ラインの特徴量が出力される。ライン毎の特徴量に対してMaxPool層でグローバル特徴量を抽出し、各ラインの特徴量に結合する。これを T 回の観測分行い、ライン毎にLSTMに入力する。LSTMからのライン毎の出力からMaxPool層で全体特徴量を抽出し、全結合層で分類を行う。

このネットワークは、ライン毎に特徴量をエンコードし分類するものである。このため、点群密度が変化してもデータの存在するラインの情報を適切に活用することができ、これまでに提案した2D-CNNベースのエンコード手法に比べて点群密度の変化によりロバストであると考えられる。

5. 実験

5.1 学習と評価方法

本実験では、作成したデータセットを用いて点群密度の変化におけるロバスト性を検証した。本実験では、4つのデータセットを学習用とテスト用にそれぞれ分割

し、4つ学習用データセットのうち3つを学習用に使い、4つのテスト用データセットで性能を確認した。学習用に用いる3つのデータセットは4パターンの組み合わせがあるため、その全てで確認を行った。なお、学習とテストは異なる周回の歩行データで分割しており、学習とテストに同じデータセットが使用されている場合でも両者に同じデータは含まれない。

5.2 実験結果

2つのネットワークによる31人の歩行者の識別結果を表1に示す。表1から、学習データに同じデータセット(点群密度)の情報が含まれる場合は高い識別率が確認でき、含まれない場合は低い識別率であることが確認できた。また、学習データに同じデータセット(点群密度)の情報が含まれない場合に着目すると、特にネットワークAにおいて、点群密度が最も高い又は最も低いデータが学習されていない場合の方が、そうでない場合に比べて相対的に識別性能が低い結果となった。これは言い換えると、未知の点群密度に対してそれよりも高い密度のデータと低い密度のデータの両方が学習されている場合は識別性能が高い傾向であると言える。

次に、ネットワークAとネットワークBを比較すると、従来のネットワークAの方が識別性能が全体的に高い結果となった。これは、データセット1~4は点群密度の違いだけではなく移動方向の違いも含まれており、ネットワークBのラインワイズの特徴量抽出では移動方向の違いによる点群の見え方の差を学習することができていないことが原因の一つと考えられる。また、ラインワイズ手法の歩容特徴では動きにおいて垂直方向の相互関係を排除していることから、識別率に悪影響を及ぼしてると考えられる。これについては、今後、移動方向が同一で点群密度が異なるデータによる性能

表 1 各距離範囲における提案手法の精度評価

Methods	Gallery				Probe			
	3.5 ~ 6m	6 ~ 9m	9 ~ 12m	12 ~ 13.5m	3.5 ~ 6m	6 ~ 9m	9 ~ 12m	12 ~ 13.5m
ネットワーク A		✓	✓	✓	0.447	0.954	0.928	0.924
	✓		✓	✓	0.957	0.887	0.967	0.951
	✓	✓		✓	0.965	0.960	0.611	0.956
	✓	✓	✓		0.964	0.938	0.930	0.501
ネットワーク B		✓	✓	✓	0.431	0.529	0.506	0.564
	✓		✓	✓	0.726	0.597	0.768	0.863
	✓	✓		✓	0.791	0.602	0.276	0.749
	✓	✓	✓		0.791	0.692	0.708	0.362

評価などを実施し、問題を切り分けた上で性能向上のためのネットワークの見直しなどを行っていく予定である。

6. 結言

本研究では、マルチライン 3D LiDAR を用いた歩容認識のため、点群密度や移動方向が異なるデータセットを構築した。また、これまでに提案したネットワークに対して、点群密度の変化にロバストであると考えられる新たなネットワークを提案した。31 人の歩行者の距離データを用いた識別実験を行い、これまでに提案したネットワークでも、歩行方向や点群密度の変化を含む歩容認識が可能であることを確認した。ただし、新たな提案ネットワークは期待した性能に達しておらず、今後のさらなる検討が必要である。

参考文献

- [1] Whytock T, Belyaev A and Robertson NM: “Dynamic Distance-Based Shape Features for Gait Recognition”, *Journal of Mathematical Imaging and Vision*, pp. 1–13, 2014.
- [2] Feng Y, Li Y and Luo J: “Learning effective Gait features using LSTM”, 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 325–330, 2016.
- [3] Yam C-Y and Nixon MS: “Model-based Gait Recognition”, *Encyclopedia of Biometrics*, pp. 633–639, 2009.
- [4] Han J and Bhanu B: “Individual recognition using gait energy image”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2005.
- [5] Gálai B and Benedek C: “Gait Recognition with Compact Lidar Sensors”, *VISIGRAPP*, pp. 426–432, 2017.
- [6] Li K, Wang X, Xu Y and Wang J: “Density Enhancement-Based Long-Range Pedestrian Detection Using 3-D Range Data”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1368–1380, 2016.
- [7] Yamada H, Ahn J, Mozos OM, Iwashita Y and Kurazume R: “Gait-based person identification using 3D LiDAR and long short-term memory deep networks”, *Advanced Robotics*, 2020, DOI:10.1080/01691864.2020.1793812.
- [8] R. Q. Charles, H. Su, M. Kaichun and L. J. Guibas, “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation”, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 77-85, doi: 10.1109/CVPR.2017.16.