

低密度 LIDAR 点群からの歩行者検出*

—3DCNN のための点群分布を考慮したボクセル表現—

建部好輝** 出口大輔*** 川西康友† 井手一郎† 村瀬 洋†

Pedestrian Detection from a Sparse LIDAR Point-Cloud
—Distribution-Based Voxel Representation for 3DCNN—

Yoshiki TATEBE, Daisuke DEGUCHI, Yasutomo KAWANISHI, Ichiro IDE and Hiroshi MURASE

In recent years, the demand for pedestrian detection using LIDAR is increasing, as it can be used to prevent traffic accidents involving pedestrians. To avoid traffic accidents, detection of distant pedestrians is very important. However, they are scanned sparsely even if a dense-scan LIDAR is used, and this causes the degradation of the detection accuracy. Therefore, pedestrian detection from sparsely-scanned LIDAR point-clouds is expected to be developed. This paper proposes a LIDAR-based pedestrian detection method using 3DCNN. Since it is difficult to train a 3DCNN directly from sparse point-clouds, the proposed method converts them to a voxel representation using the kernel density estimation based on LIDAR characteristics. To evaluate the performance of the proposed method, an experiment using real-world LIDAR data was conducted. The results showed that the proposed method could detect pedestrians more accurately than detectors trained with other conventional features.

Key words: LIDAR, 3DCNN, pedestrian detection, voxel representation, sparse point-cloud

1. 緒 言

近年、自動車の自動運転技術や先進安全運転支援システムへの注目が高まっており、実用化に向けた取り組みも盛んに行われている。自動運転車両を実現するためには、自動車の周囲環境認識技術は必要不可欠であり、その中でも歩行者を認識する技術は特に重要となる。車載カメラ画像を対象とした歩行者検出技術が広く研究されてきた一方で、ここ数年は実用化の観点から距離センサを併用した技術への注目も高まっている。その中でも、LIDAR (LIght Detection And Ranging) は大きな注目を集めており、大衆車への搭載を目指した安価な製品の開発も進んでいる。図 1 は安価な距離センサで取得したデータの例を示している。図 1 から分かるように、安価な LIDAR から得られる 3 次元データは非常に疎 (低密度) であり、歩行者と類似した静止物 (電柱など) の識別が難しいというような課題がある。

これまでに、LIDAR を対象とした歩行者検出の研究がいくつかなされており、LIDAR により観測される 3 次元点群データから抽出可能な特徴量が提案されている¹⁾²⁾。Kidono らは、中～遠距離の歩行者の検出を目的として、歩行者の大きな形状を捉えるスライス特徴量と反射強度の分布に関する特徴量を提案している¹⁾。また我々は、時系列情報を考慮することで点群の密度を高める特徴抽出手法、点群の時間変化を捉える手法を提案している²⁾。これらを組み合わせて用いることにより、遠方で比較的低密度な点群として観測された歩行者の検出精度向上に成功している。しかし、より遠方で非常に密度が低くな

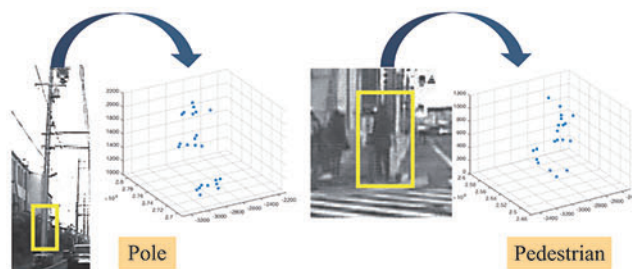


Fig. 1 Examples of sparse LIDAR point-clouds

る場合や歩行者と類似する電柱等の物体の識別精度は依然として低いという問題があった。

一方、近年画像認識分野を中心に深層学習手法が急速に発展してきており、3 次元点群を対象とした物体認識の研究も行われつつある³⁾⁴⁾。Maturana らは、3 次元点群の分布をボクセル空間で表現することにより、3 次元点群からの物体認識に深層学習を適用する手法を提案している⁴⁾。この手法では、占有領域、非占有領域、未知領域の 3 値からなる占有グリッドによってボクセル空間を構成し、3 次元点群の分布を表現している。そして、ボクセル表現されたデータを 3D Convolutional Neural Network (3DCNN) に入力することにより、鉛直方向の点群の分布を明示的に考慮した認識を可能としている。その結果、様々な物体を高精度で識別することに成功している。しかしながら、この手法は CAD や高密度な点群を取得可能な LIDAR (高密度 LIDAR) から得られる非常に高密度な点群の入力が前提であり、低密度な点群しか取得できない安価な LIDAR (低密度 LIDAR) から得られる極めて疎な点群には直接適用できない。具体的には、低密度な点群をボクセル表現に

* 原稿受付 平成 30 年 5 月 7 日
掲載決定 平成 30 年 7 月 13 日

** 名古屋大学大学院情報科学研究科 (愛知県名古屋市千種区不老町)

*** 名古屋大学情報連携統括本部 (愛知県名古屋市千種区不老町)

† 名古屋大学大学院情報科学研究科 (愛知県名古屋市千種区不老町)

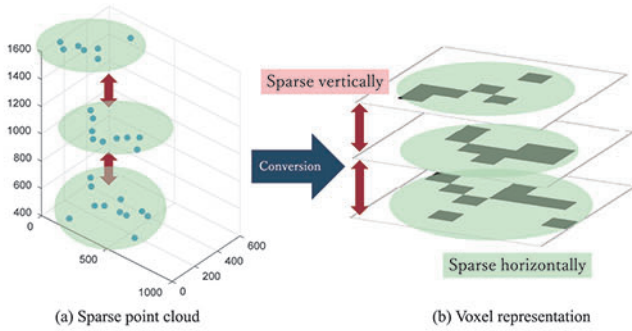


Fig. 2 Example of conversion to voxel representation from a sparse point cloud ($L = 3$)

変換した場合、図 2 に示すように鉛直、水平両方向において疎な表現となってしまう、3DCNN の学習を適切に行えないという問題がある。

そこで本論文では LIDAR の特性を考慮することにより、低密度点群を 3DCNN で認識可能なボクセル表現へと変換する手法を提案する。本論文で対象とする低密度 LIDAR は、鉛直方向の走査線数 L ($L = 3 \sim 6$) が非常に少ないことから、得られる 3 次元データは鉛直方向に非常に疎である。しかし、歩行者は柱状の物体であることから、鉛直方向の走査位置が多少上下に変化したとしても、走査線という単位で見た際の形状は大きく変化しない。そこで、まず各水平走査により得られた点群に分割し、鉛直方向が走査線に対応するボクセル表現へと変換する。また LIDAR の誤差は、対象までの距離が遠くなるに従って大きくなるというセンサ特性が存在する。そこでこの特性を考慮することにより、近くと遠くの点群の計測誤差を考慮した上でボクセル空間に変換する手法を提案する。更に、Maturana⁴⁾ の手法においては考慮されていない反射強度の情報をボクセル空間上に取り込み、3DCNN での認識に利用する方法を提案する。

以上をまとめると、本論文の貢献は以下の 3 点に集約^{*1}される。

1. 低密度 LIDAR から得られる疎な点群からの歩行者検出。
2. LIDAR の特性を考慮した計測点群のボクセル空間表現。
3. 反射強度を考慮したボクセル空間表現。

以降、2. で低密度 LIDAR を用いた歩行者検出手法について述べ、3. で提案手法の評価実験について述べる。そして、4. で結果を考察し、最後に 5. で本論文をむすぶ。

2. 低密度 LIDAR を用いた歩行者検出

前節で述べたように、低密度 LIDAR の鉛直方向の走査線数 L ($L = 3 \sim 6$) は非常に少ない。しかし、歩行者は柱状の物体であることから、車の振動等によって水平走査線の高さが多少変化したとしても各走査線から得られる形状は大きく変化しない。そこで、ボクセル表現の鉛直方向を LIDAR の各走査線と対応付けることにより、3 次元点群の鉛直方向が疎である問題に対処する。一方、LIDAR の LASER 光は一定間隔で照射され、LASER 光の照射間隔とその太さ (幅) によって定められる範囲内から観測点を得られる。この様子を図 3 に示す。図 3

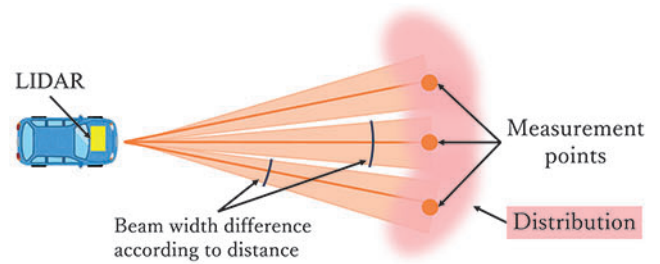


Fig. 3 Illustration of LIDAR's characteristics

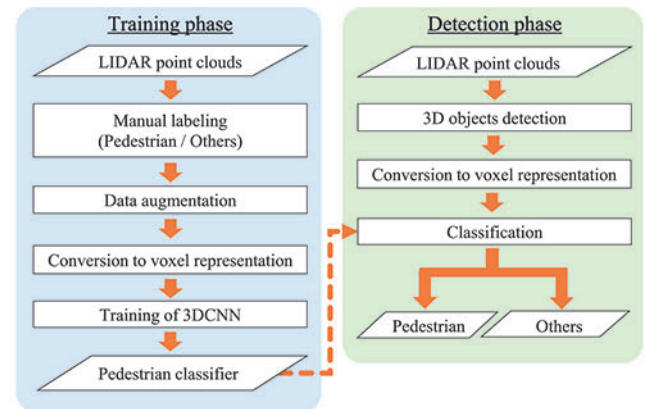


Fig. 4 Process flow of the proposed method

から分かるように、観測結果は LASER 光の間隔分の幅の曖昧性をもって得られる。そのため、観測点は 3 次元上のある特定の 1 点ではなく、ある幅をもった確率分布のように扱うことができると考えられる。また、その幅は対象物体までの距離が遠くなるほど大きくなる。そこで、距離に応じた観測点の曖昧性を考慮することにより、近くと遠くの点群を同じ枠組みでボクセル空間へと変換する。更に、LIDAR から距離に加えて対象の反射強度も同時に計測が可能である。そこで、距離情報と反射強度情報を同じボクセル空間上で表現し、2 種類のチャンネルを入力とする 3DCNN によって歩行者を識別する手法を実現する。

提案手法の処理手順を図 4 に示す。提案手法は、事前に歩行者検出に使用する識別器を構築する学習段階と、観測した LIDAR データから歩行者を検出する検出段階の 2 段階に分かれる。

学習段階では、学習用の点群を用いて歩行者識別器を構築する。まず、学習用の点群として、歩行者点群 (Positive Sample) と非歩行者点群 (Negative Sample) を用意する。これらの学習用点群は、実環境で計測した LIDAR データに対して立体物検出・追跡を行い、得られた各立体物に対して人手で歩行者か非歩行者かのラベル付けをすることで得られる。次に、各点群からデータを疑似的に生成し、ボクセル表現を求める。最後に、ボクセル表現を 3DCNN に入力することで歩行者と非歩行者を識別する識別器を構築する。

検出段階では、LIDAR から得られる 3 次元点群データに対して、学習段階と同様に立体物検出と追跡アルゴリズムを適用し、立体物点群を抽出して候補点群とする。次に、候補点群系列から学習段階と同様にボクセル表現に変換する。最後に、変換後のボクセル表現を学習段階で構築した識別器に入力することで、歩行者候補点群が歩行者であるか否かを識別し、検出結

*1 本論文は文献⁵⁾を発展させてまとめたものである。

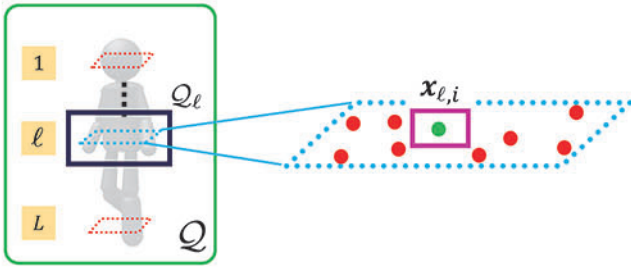


Fig. 5 Illustration of the structure of a point cloud candidate

果とする。

学習段階と検出段階はほぼ同じ処理で構成されることから、以降では学習段階の各手順について詳細に説明する。

2.1 立体物検出

提案手法では、自動車の周囲を観測した LIDAR データから立体物点群を Ogawa らが提案したアルゴリズム⁶⁾で抽出する。以降、立体物検出で得られた点群を歩行者候補点群と呼ぶ。

本論文では、立体物をとらえた水平スキャンの数を構成ライン数 L とし、計測したライン毎に歩行者候補点群を分割する。ここで、歩行者候補点群は次のような階層的表現によって表すことができる。

$$\begin{aligned} Q &= \{Q_\ell\}_{\ell=1}^L, \\ Q_\ell &= \{\mathbf{x}_{\ell,i}\}_{i=1}^{N_\ell}, \end{aligned} \quad (1)$$

Q はある時刻で観測した候補内の全ての点群であり、 Q_ℓ は ℓ 番目の水平スキャンによって得られた N_ℓ 個の計測点、 $\mathbf{x}_{\ell,i}$ は Q_ℓ の内の i 番目の計測点を表す (図 5 参照)。また、 $\mathbf{x}_{\ell,i}$ に対して計測された LASER 光の反射強度は $r_{\ell,i} = I(\mathbf{x}_{\ell,i})$ と表す。

さらにボクセル表現へ変換する事前処理として、各フレームの点群の向きと位置を正規化する。具体的には、歩行者候補点群 Q を道路面に射影し、主成分分析 (PCA) を施す。次に、PCA によって得られたフレーム毎の主軸を座標軸とし、各点群の重心が原点となるように点群を平行移動する。この事前処理を施した点群を以下の処理の入力とする。

2.2 疎な点群からボクセル表現への変換

提案手法では、立体物の存在尤度と反射強度からなる 2 チャネルボクセル表現 $\mathcal{V}^d, \mathcal{V}^r$ を 3DCNN の入力として用い、候補点群 $Q(t)$ からこれらを求める。図 6 に $L=3$ の場合のボクセル表現への変換手順を示す。この処理は、大きく 3 段階に分かれる。第 1 段階では、各候補点群 Q を各水平走査により得られた部分集合 Q_ℓ ($1 \leq \ell \leq L$) へ分割する。第 2 段階では、分割後の各点群 Q_ℓ 毎に存在尤度マップ M_ℓ^d と反射強度マップ M_ℓ^r を求める。第 3 段階では、各特徴マップを 2 チャネルボクセル表現 \mathcal{V}^d と \mathcal{V}^r に変換する。

2.3 点群の分割

各候補点群を各水平走査により得られた点群 Q_ℓ へ分割してボクセル表現への変換に利用する。 Q_ℓ 毎に処理することにより、ボクセル空間において鉛直方向で観測できていない未知領域を考慮する必要がなくなるだけでなく、水平走査が当たる高さが多少ずれたとしても、構成ライン数が同じであればほぼ同一のボクセル表現を得ることができる。このようにして低密度 LIDAR から得られる 3 次元点群が鉛直方向において疎な問題

に対処する。

2.4 特徴マップの生成

本手法で扱う低密度な点群データは、遠方の歩行者や障害物から抽出したものが多く、水平方向にも疎であるために、直接ボクセル表現の算出に利用するのは難しい。そこで、各観測点の距離と反射強度に基づいたカーネル密度推定を利用して、特徴マップを得る。

本手法では 2 種類の特徴マップを用いる。1 つ目は存在尤度マップであり、各ボクセルはその位置に対象物体が存在する尤度を示す。2 つ目は反射強度マップであり、各ボクセルはその位置における対象物体の反射強度を示す。以降、各マップの生成手順について説明する。存在尤度マップ M_ℓ^d は、各観測点 $\mathbf{x}_{\ell,i}$ を用いて、

$$\sum_{i=1}^{N_\ell} g(\mathbf{P}\mathbf{x}_{\ell,i}, \Sigma_{\ell,i}) \quad (2)$$

のように表される。ここで、 g はカーネル関数である 2 次元正規分布、 \mathbf{P} は道路面への射影行列である。つまり、道路面に射影した点を中心とし、観測点毎に異なる分散共分散行列 $\Sigma_{\ell,i}$ をもつ 2 次元混合正規分布 $g(\mathbf{P}\mathbf{x}_{\ell,i}, \Sigma_{\ell,i})$ として存在尤度マップを得る。ここで、分散・共分散行列 $\Sigma_{\ell,i}$ は各観測点までの距離 $d_{\ell,i} = \|\mathbf{x}_{\ell,i}\|$ に依存し、次式のように表される。

$$\Sigma_{\ell,i} = \begin{pmatrix} \sigma_{\ell,i}^x & 0 \\ 0 & \sigma_{\ell,i}^y \end{pmatrix}, \quad (3)$$

$$\sigma_{\ell,i}^x = d_{\ell,i} \tan \alpha, \quad (4)$$

$$\sigma_{\ell,i}^y = E(d_{\ell,i}) \quad (5)$$

ここで、 $\sigma_{\ell,i}^x$ と $\sigma_{\ell,i}^y$ は、いずれも $d_{\ell,i}$ と LIDAR の計測誤差に基づいて計算される。具体的には、 $\sigma_{\ell,i}^x$ は水平誤差に対応し、ビーム幅と計測誤差の 2 つのセンサ特性から定まる。また、 $\sigma_{\ell,i}^y$ は奥行誤差に対応し、センサ特性に基づいて定まる値である。このようにカーネル関数を決定することで、図 7 に示すように、遠方の観測点に対しては大きなカーネル、近辺の観測点に対しては小さなカーネルを当てはめることができる。

一方、反射強度マップ M_ℓ^r は距離による減衰を補正した正規化反射強度 $r_{\ell,i}^*$ を利用し、

$$\sum_{i=1}^{N_\ell} r_{\ell,i}^* \cdot g(\mathbf{P}\mathbf{x}_{\ell,i}, \Sigma_{\ell,i}) \quad (6)$$

のように表される。ただし、

$$r_{\ell,i}^* = r_{\ell,i} \cdot d_{\ell,i}^2 \quad (7)$$

である。図 8 に各チャネルの特徴マップの例を示す。

2.5 ボクセル表現への変換

2 チャネルボクセル表現 \mathcal{V}^d と \mathcal{V}^r は、上述の 2 種類の特徴マップを一定間隔で標本化することで得られる。具体的には、次式によって各ボクセルの値を求める。

$$\mathcal{V}(j, k, \ell) = \int_{\Omega_{j,k}} M_\ell(\mathbf{x}) d\mathbf{x} \quad (8)$$

ここで積分領域 $\Omega_{j,k}$ は、次のように定義される。

$$\Omega_{j,k} = \left\{ \bar{\mathbf{x}} + \Delta\mathbf{x} \mid \gamma \|\Delta\mathbf{x} - (j, k)^T\|_\infty < 0.5, \Delta\mathbf{x} \in \mathbb{R}^2 \right\} \quad (9)$$

ここで、 $\|\cdot\|_\infty$ は ∞ ノルムを表し、また、

$$\bar{\mathbf{x}} = \frac{1}{|Q|} \sum_{\mathbf{x} \in Q} \mathbf{P}\mathbf{x}, \quad (10)$$

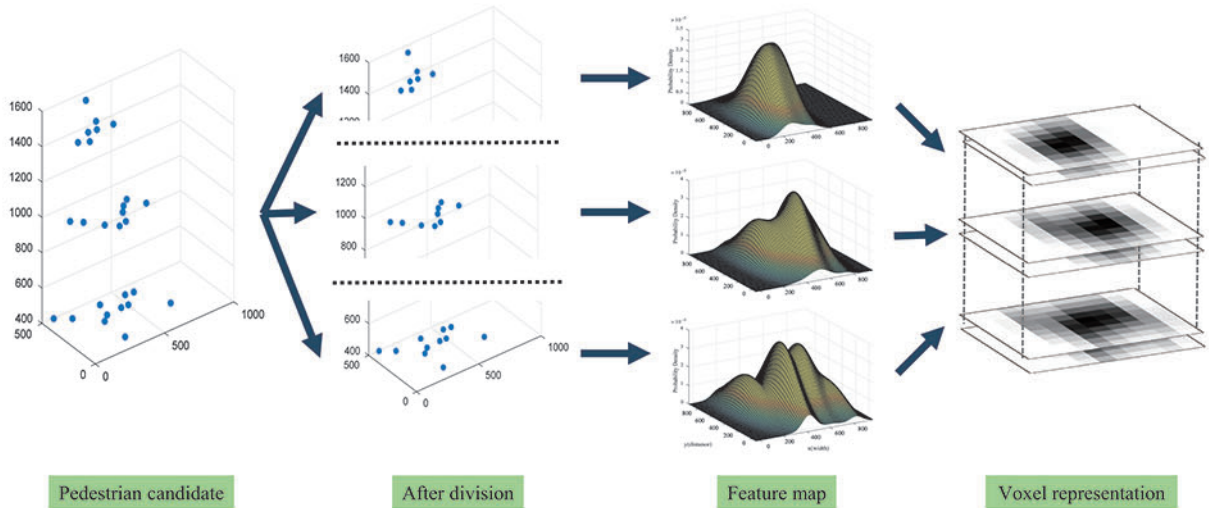


Fig. 6 Conversion to voxel representation ($L = 3$). The unit length of each graph is in mm

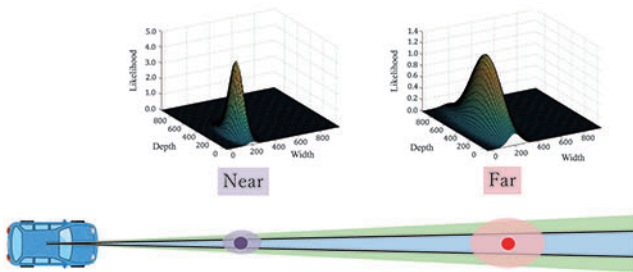


Fig. 7 Shape of error distribution according to the distance

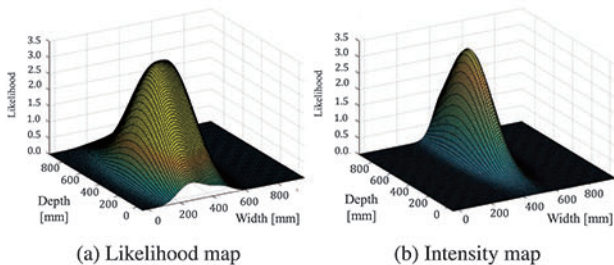


Fig. 8 Example of feature map in each channel

である。ただし、 \mathbf{P} は道路面への射影行列であるため、 $\bar{\mathbf{x}} \in \mathbb{R}^2$ である。 γ はボクセル間隔に対応する。ここでは、 $\gamma = 0.1$ と設定し、10cm 四方で標本化する。また、 $\bar{\mathbf{x}}$ を中心として特徴マップの 1m 四方の範囲を標本化することで、最終的にボクセル表現は $10 \times 10 \times L \times 2 = 200L$ 次元となる。

2.6 学習用点群のデータ拡張

3DCNN の学習に用いるデータを増やすために、学習用の点群に摂動と回転を加え、学習標本の数を 10 倍に増やす。ここで、摂動の量は LIDAR の測定誤差を考慮して設定し、回転は $[-10^\circ, 10^\circ]$ の範囲で一様に点群を回転させる。この摂動と回転を組み合わせることにより、多様な学習パターンを生成する。

Table 1 Number of pedestrians/others in scan lines

	Number of scan line			
	$L = 3$	$L = 4$	$L = 5$	$L = 6$
Pedestrian	1,134	5,343	7,016	8,258
Others	7,143	7,360	8,477	14,719

Table 2 Average distance in meters of pedestrians/others in scan lines. Standard deviation is shown in the bracket

	Number of scan line			
	$L = 3$	$L = 4$	$L = 5$	$L = 6$
Pedestrian	27.5 (5.3)	25.4 (4.7)	18.5 (4.0)	13.0 (3.8)
Others	21.7 (5.9)	20.5 (6.5)	21.4 (5.5)	14.8 (4.6)

2.7 識別器の構築

本論文では識別器として 3DCNN (3D Convolutional Neural Network) を用いる。3DCNN は、画像処理で広く用いられている CNN (Convolutional Neural Network) の畳み込み層を 3 次元に拡張したモデルであり、拡張した 3 次元目を時間軸ととらえて行動認識問題によく用いられている。3DCNN の入力の前節までに求めたボクセル表現であり、図 9 のように構成ライン数 L 毎に学習する。 L によって入力の次元数は変わるが、ネットワーク構造は L によらず共通であり、以下の 4 層からなる。

- 3D 畳み込み層 ($32, 3 \times 3 \times 2, 1$)
- 3D 畳み込み層 ($32, 3 \times 3 \times 2, 1$)
- 全結合層 (128)
- 全結合層 (2)

3D 畳み込み層 (f, k, s) は f 種のフィルタ (大きさは k) でストライド s での畳み込みを表し、全結合層 (n) は n 出力の全結合層である。活性化関数には ReLU (Rectified Linear Units)⁷⁾ を用いる。

3. 実験

本論文で提案する深層学習を活用した歩行者識別手法の有効性を確認するために行った評価実験について述べる。まず、3.1 で評価実験における実験条件について説明する。次に、3.2 で評価実験の結果について報告する。

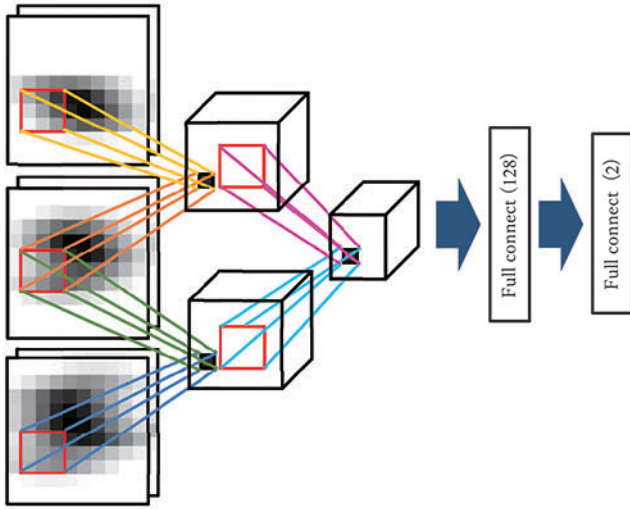


Fig. 9 Structure of 3DCNN ($L = 3$)

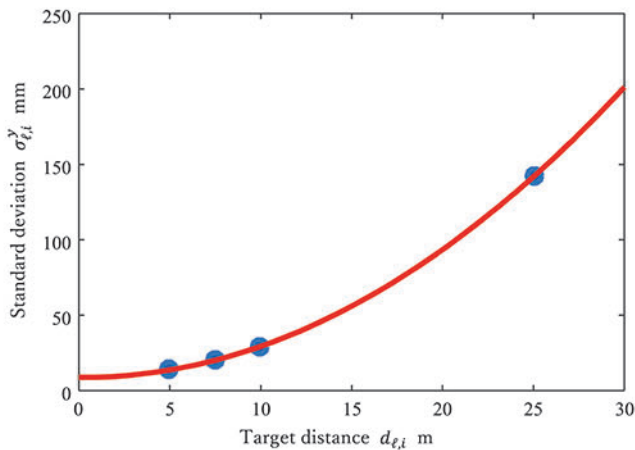


Fig. 10 Result of quadratic fit for LIDAR measurement errors

3.1 実験条件

本実験では走査線数が6本(最大で $L = 6$)の低密度 LIDAR を用い、実環境でデータ収集を行った。実験で用いた低密度 LIDAR は自動車のルームミラー上部に搭載されており、6本の走査線で一定間隔に $401 \times 6 = 2,406$ 点を10fpsで計測ができる。得られたデータに対してOgawaらの立体物検出・追跡アルゴリズム⁶⁾を適用し、抽出された立体物の点群データを用いて識別機の性能を評価した。シーン中にはOgawaらの立体物検出・追跡アルゴリズムで抽出できない歩行者が存在する可能性はあるものの、LIDAR上の3次元点群に対してそれらすべてに対するアノテーションを手手で付与することは困難である。そのため、本実験ではOgawaらの手法で抽出した立体物に対して手手で歩行者かどうかのラベルを付与して性能評価に用いた。手作業でアノテーションを付与した実験用データを構成ライン数別に集計した結果を表1に示す。また、各対象までの距離と標準偏差を構成ライン数別に集計した結果を表2に示す。歩行者候補点群を構成ライン数別に分割し、それぞれに対して2分割交差検証法によって性能を評価した。また、データセットを分割する際に学習データと評価データに同じ物体から得られた点群が含まれないようにした。評価指標とし

Table 3 Measurement error of LIDAR used in the experiment

Groundtruth distance in meters	5.000	7.500	10.000	25.000
Ave. of measurement in meters	4.995	7.569	10.037	25.070
Std. of measurement in meters	0.014	0.020	0.029	0.142

Table 4 Evaluation methods

	Conventional	Comparative 1	Comparative 2	Proposed
Feature	Handcrafted ¹⁾	Voxel	Voxel	Voxel
Classifier	SVM	SVM	2DCNN	3DCNN

Table 5 Results of AUCs (Highest scores are indicated in bold font)

Method	Number of scan line			
	$L = 3$	$L = 4$	$L = 5$	$L = 6$
Conventional	0.924	0.931	0.975	0.974
Comparative 1	0.915	0.931	0.966	0.967
Comparative 2	0.915	0.953	0.970	0.980
Proposed	0.921	0.967	0.981	0.984

て、ROC (Receiver Operating Characteristic) 曲線と AUC (Area Under the Curve), p-AUC⁸⁾ (partial Area Under the Curve) を用いた。p-AUCとは、ROC曲線のある区間のみから求めたAUCであり、最大値が1になるように正規化されている。p-AUCをROC曲線のうち誤検出率(FPR)が10%以下の区間でp-AUCを求めることで、歩行者検出にとって重要となるFPRが低い区間に着目して性能を評価する。

また実験に用いたセンサ特性に基づき、式(4)には $\alpha = 0.1^\circ$ を用いた。式(5)の $\sigma_{\ell,i}^y$ については、複数の距離でセンサの奥行の測定誤差を測定し(表3参照)、その結果に対して2次曲線を当てはめた結果(図10参照)を用いた。

一方2.6のデータ拡張では、図10に示す $\sigma_{\ell,i}^y$ を水平、鉛直両方向における標準偏差とした2次元ガウス分布より求めた正規乱数を摂動として加えるだけでなく、一緒に回転させることで学習用の点群数を増やした。具体的には、1つの点群から摂動のみで3つ、回転のみで3つ、摂動と回転の組み合わせで3つ点群を生成し、データ数を10倍に増やした。また、2.7における学習パラメータとして、バッチサイズ128、エポック数100を用いた。

提案手法の有効性を確認するため、以下に示す従来手法、比較手法1、比較手法2、提案手法の4つの手法を比較した。従来手法は、従来の人手で設計したハンドメイド特徴量¹⁾⁹⁾をカーネルSVMによって識別する手法である。比較手法1は、本論文で提案するボクセル表現を特徴量とみなし、従来手法と同じカーネルSVMにより識別する手法である。ここで、従来手法と比較手法1のいずれの手法も、SVMのカーネルとしてRBFカーネルを用いた。比較手法2と提案手法は共にCNNを識別器として用いる手法である。比較手法2と提案手法のいずれの手法も2.7で述べたネットワーク構造と同様のCNNを用いたが、比較手法2では3次元畳み込み層を通常の2次元畳み込み層に置き換えた手法である。表4に各手法の特徴を示す。

3.2 実験結果

図11に実験により得られた各手法の構成ライン数別のROC曲線(誤検出率10%までを描画)を示す。提案手法は従来手法、比較手法1、比較手法2と比べて検出率(TPR)が上回っており、歩行者検出において重要である低FPR時の識別精度が向上したことが分かる。また、表5に実験により得られた

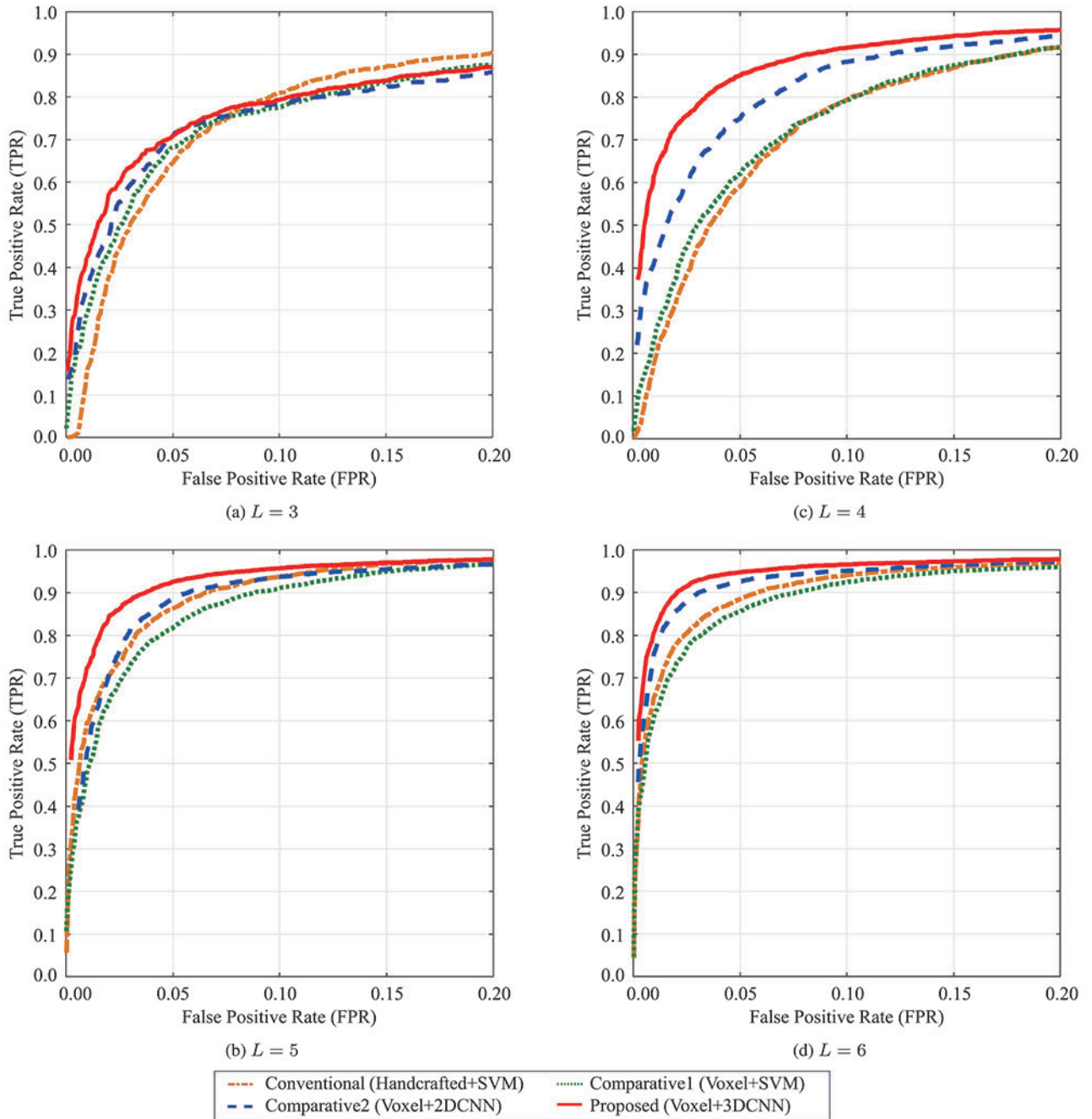


Fig. 11 ROC curves of the results

Table 6 Results of p-AUC (Highest scores are indicated in bold font)

Method	Number of scan line			
	$L = 3$	$L = 4$	$L = 5$	$L = 6$
Conventional	0.564	0.535	0.805	0.836
Comparative 1	0.601	0.565	0.757	0.807
Comparative 2	0.629	0.698	0.809	0.884
Proposed	0.658	0.802	0.881	0.914

各手法の構成ライン数別の AUC を示す。提案手法では $L = 3$ の場合を除くと、従来手法、比較手法 1、比較手法 2 よりも高い AUC を得られていることが分かる。一方、各手法の FPR が 10% 以下の区間のみで求めた p-AUC を表 6 に示す。提案手法では従来手法、比較手法 1、比較手法 2 よりも高い p-AUC を

得られていることがわかる。そのため、提案手法では歩行者検出に重要な低 FPR 時における識別精度の向上が顕著であることを確認した。

また、その他の物体としては Ogawa らの立体物検出・追跡アルゴリズムによって抽出された歩行者と同程度の大きさを持つ物体が含まれており、誤検出としては歩行者と類似した形状を持つ電柱、ポール、樹木の一部等が多く見られた。

4. 考 察

4.1 3次元畳み込みの効果

比較手法 2 と提案手法に注目すると、3次元畳み込みを利用する提案手法の精度が高いことが分かる。3次元畳み込みによって、3次元点群の3次元的な分布を考慮することができる。

識別が困難な木や電柱は、歩行者と似た円柱上の形状である。そのため、ある高さでの形状しか考慮しない2次元畳み込みでは識別が難しい。一方、鉛直方向の点群の分布を考えると、木や電柱はどの高さで切り出しても変化が小さいのに対して、歩行者は比較的变化が大きい。そのため、3次元データからの歩行者検出には点群の3次元構造を考慮することが有効であったと考えられる。

表2から分かるように、対象の構成ライン数が増加するにつれて対象までの距離が近くなる。そこで対象までの距離による性能の変化という観点で表5と表6を見ると、距離が遠くなる（構成ライン数が減少する）に従って提案手法の性能が低下していることが確認できる。この傾向は、図11からも確認することができる。特に構成ライン数が $L=3$ において性能が大きく低下しており、非常に遠方の対象に対しては3次元畳み込みの効果も限定的であることが分かる。今後、時系列情報の活用等によるさらなる性能向上が必要だと考えられる。

4.2 ボクセル表現の表現力

比較手法1と従来手法に注目すると、ボクセル表現を特徴量としてSVMで識別するだけでも従来の特徴量と同等の性能が得られている。これは、提案したボクセル表現には各従来特徴量を求めるために必要な情報が含まれているためであると考えられる。具体的には、形状特徴であるスライス特徴量は2チャンネルボクセル表現のうち存在尤度チャンネルの分布から算出できる。反射特性には、2チャンネルボクセル表現のうち、反射強度チャンネルが相当する。このように定量化された各チャンネルのボクセル表現からSVMが自動的に従来の特徴量と同等の特徴を学習したと考えられる。

4.3 ネットワーク構造による識別性能の変化

低密度でかつデータセット中で正例と負例の数のバランスが良い構成ライン数 $L=4$ を対象として、ネットワーク構造による認識精度の違いを分析した結果について述べる。具体的には、提案手法で用いたネットワーク構造(3D-3*3*2)に加え、以下に示す2種類のネットワーク構造を比較した。

1. 3D-3*3*3

- 3D畳み込み層(32, 3 × 3 × 3, 1)
- 全結合層(128)
- 全結合層(2)

2. 3D-3*3*2+2D-3*3

- 3D畳み込み層(32, 3 × 3 × 2, 1)
- 3D畳み込み層(32, 3 × 3 × 2, 1)
- 2D畳み込み層(32, 3 × 3, 1)
- 2D畳み込み層(32, 3 × 3, 1)
- 全結合層(128)
- 全結合層(2)

比較に用いた $L=4$ のデータ数は表1の通りである。また、同一系列の点群が同じ分割内に含まれないように2分割し、各々を全てのネットワークで学習データと評価データとして用いた。

各ネットワークにおけるROC曲線を図12に示す。誤検出率(FPR)が10%における各ネットワークの正検出率(TPR)はほぼ同様であるが、全体的に提案手法で用いた3D-3*3*2が最も良い結果であることが分かる。しかし、各ネットワーク間の性能差はわずかである。この結果は、本論文で提案したボク

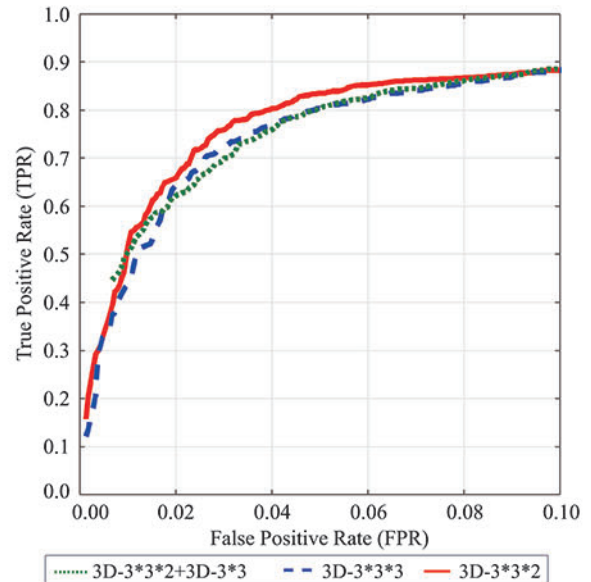


Fig. 12 ROC curves of the results for different network structures

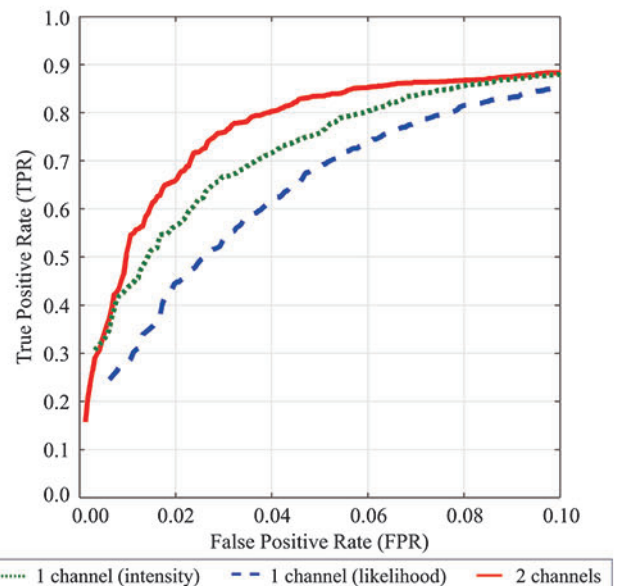


Fig. 13 ROC curves of the results for different input channels

セル表現の次元数が非常に小さいため、非常に小規模なネットワークによって十分に学習および識別が可能であることを示唆している。

4.4 複数チャンネル同時利用の効果

4.3と同様に構成ライン数 $L=4$ のデータを用い、本論文で提案する反射強度チャンネルの有効性について考察する。具体的には、提案手法である2チャンネル(存在尤度+反射強度)と存在尤度チャンネルのみを利用する1チャンネル(存在尤度)、そして反射強度チャンネルのみを利用する1チャンネル(反射強度)を比較した。実験条件は4.3と同じである。

各手法におけるROC曲線を図13に示す。本論文で提案する反射強度チャンネルを用いることで、低誤検出率時の正検出率が向上していることが分かる。まず、1チャンネル(存在尤度)と1チャンネル(反射強度)を比較すると、反射強度を考慮する1チャンネル(反射強度)が高精度である。この結果から、歩行者を識別する際に反射強度が有効であることが分かる。また、

2チャンネルと1チャンネルのみの2手法を比較すると、存在尤度チャンネルと反射強度チャンネルを同時に学習に利用する2チャンネルが高精度である。点群の距離情報のみから求めた存在尤度チャンネルからは物体形状に関する特徴しか抽出できないのに対し、反射強度チャンネルも学習に用いることで、形状とその物体の反射特性を併せて学習することが出来る。具体的には、計測対象の形状とその反射強度が強い箇所等を学習できる。そのため、形状が類似した電柱や樹木と歩行者を区別する場合に、一様な材質で構成された電柱や樹木と衣類等で多様な材質によって構成された歩行者を区別できるようになったと考えられる。これらの理由から、歩行者を高精度に識別するためには歩行者の形状と反射特性を合わせて学習することが有効であると考えられる。

5. 結 言

本論文では、3DCNNを用いた低密度LIDARによる歩行者検出手法を提案した。具体的には、低解像度な歩行者を精度良く検出するために、各点の距離や反射強度を利用した尤度マップを構築し、低解像度な点群をボクセル表現へと変換した。そして、LIDARにより計測されたデータを構成ライン数別に分割して識別器を学習し、歩行者検出を行った。実環境で収集したデータを用いた実験により、3DCNN+提案ボクセル表現は既存の特徴量よりも高い精度で識別が可能であることを確認した。また、3次元畳み込みの有効性も確認し、歩行者検出には鉛直方向の点群の分布を考慮することが有効であることを確認した。今後の課題としては、時系列情報を利用してボクセル表現を求める、あるいは立体物検出手法の改良などが挙げられる。

謝 辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究の一部は、科学研究費補助金による。

参 考 文 献

- 1) K. Kidono, T. Miyasaka, A. Watanabe, T. Naito and J. Miura: Pedestrian recognition using high-definition LIDAR, Proceedings of 2011 IEEE Intelligent Vehicles Symposium, (2011) 405.
- 2) Y. Tatebe, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase and U. Sakai: Can we detect pedestrians using low-resolution LIDAR?, Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, (2017) 157.
- 3) Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao: 3D shapenets: A deep representation for volumetric shapes, Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, (2015) 1912.
- 4) D. Maturana and S. Scherer: Voxnet: A 3D convolutional neural network for real-time object recognition, Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, (2015) 922.
- 5) Y. Tatebe, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase and U. Sakai: Pedestrian detection from sparse point-cloud using 3DCNN, Proceedings of International Workshop on Advanced Image Technology, C4-4 (2017) 1.
- 6) T. Ogawa, H. Sakai, Y. Suzuki, K. Takagi and K. Morikawa: Pedestrian detection and tracking using in-vehicle LIDAR for automotive application, Proceedings of 2011 IEEE Intelligent Vehicles Symposium, (2011) 734.
- 7) V. Nair and G.E. Hinton: Rectified linear units improve restricted Boltzmann machines, Proceedings of the 27th International Conference on Machine Learning, (2010) 807.
- 8) S. Walter: The partial area under the summary ROC curve, Statistics in Medicine, **24**, 13, (2005) 2025.
- 9) 市川善規, 出口大輔, 井手一郎, 村瀬洋, 三澤秀明, 酒井映: 低解像度LIDARを用いた歩行者検出のための特徴量に関する検討, 電子情報通信学会技術研究報告, PRMU2014-55, (2014) 7.