

列車前方支障物検知のためのカメラと LiDAR センサの統合手法の検討

○ [電] 影山 椋 [電] 長峯 望 [電] 向嶋 宏記 (鉄道総合技術研究所)

Study on sensor fusion method between camera and LiDAR
for train frontal obstacle detection

○Ryo Kageyama, Nozomi Nagamine, Hiroki Mukojima (Railway Technical Research Institute)

Recently, the importance of robust obstacle detection method for railways is increasing as with those of automobiles. Therefore, we have developed a method to integrate camera image and point cloud data from LiDAR for obstacle detection in front of a train. To complement the detection performance of camera alone, the point cloud data from LiDAR is superimposed on the objectness distribution generated by detection algorithm from the camera images. In this paper, we evaluate the accuracy of our method using nighttime image and point cloud data.

キーワード：カメラ，LiDAR，センサフュージョン，列車前方支障物検知

Key Words：Camera, LiDAR, Sensor fusion, Train frontal obstacle detection

1. はじめに

鉄道の線路内における支障物との衝突事故を防止することは、列車の安全・安定運行を維持するために重要である。自動車業界では昨今、カメラやセンサを用いて周囲の支障物を自動的に検知する技術の開発が進められているが、鉄道でも同様に、目視を補完する前方監視技術が求められている。鉄輪で鉄軌道上を走行する鉄道の場合、自動車と比較して車輪と路面間の摩擦係数が小さいため、制動距離が長い。したがって自動車で応用されている検知手法と比べてより遠方の、小さく見える対象物も検知する必要がある。

そこで我々は、列車前方に設置したカメラで撮影した前方映像から、物体検知アルゴリズムで線路内の支障物を検知する手法を開発してきた¹⁾。昼夜にわたる人物を被写体とした評価試験の結果、提案手法では昼間に列車の300m前方に立つ人物に対して90%以上の確率で検知できることを確認した。その一方で、画像中で対象物が不明瞭になる夜間は検知率が低下することが分かった。そこで本研究では、LiDAR センサから点群データを取得することで、夜間におけるカメラの検知性能の低下を補完するための統合手法を提案する。

本発表では最初に、本研究における統合手法について、取得された画像および点群データから物体の候補領域の情報を求め、それらを統合する手法について述べる。次に、本手法の精度評価のために行った実験の概要、およびその結果について述べる。

2. カメラと LiDAR の統合手法

2.1 提案手法の全体像

提案手法の全体像を図1に示す。本手法ではカメラで画像を取得し、深層学習による物体検出アルゴリズムで物体の分類と存在確率を算出すると同時に、LiDAR で物体にレーザー光を照射して跳ね返ってきた位置の距離情報と反射率を点群データとして取得する。画像に点群データを投影し物体検出アルゴリズムで得られた物体候補領域の情報と点群の情報を組み合わせて最終的な検知判定を行う。

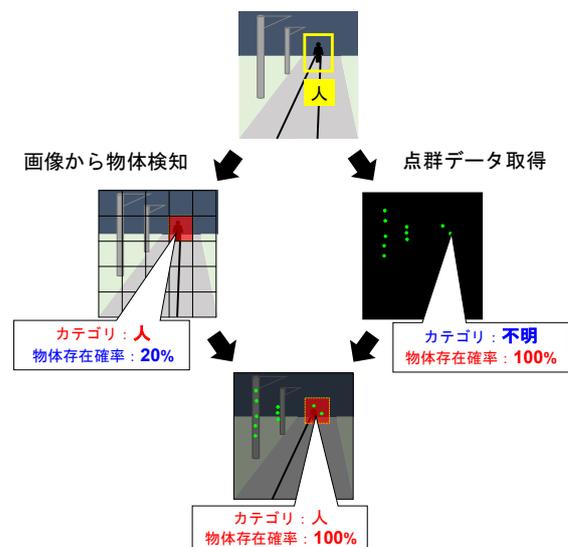


図1 本手法の全体像

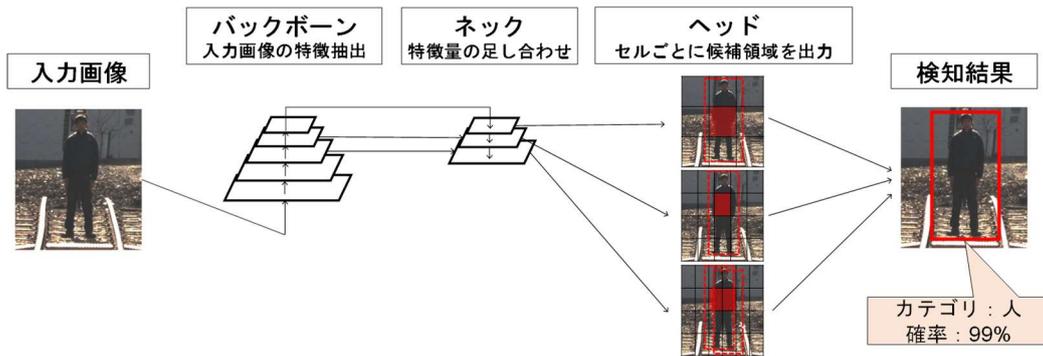


図2 YOLOv4による物体検知

2.2 画像からの物体検知

前方面像から物体検知を行う手法として、深層学習手法である YOLOv4²⁾を用いる。YOLOv4 による物体検知の流れを図2に示す。最初に入力画像の各画素は、バックボーンと呼ばれるニューラルネットワークで複数回畳み込まれる。畳み込み後の特徴量は、畳み込みの回数が多い(深い)層ほど画像中の大域的な特徴を、反対に少ない(浅い)層ほど局所的な特徴を表現する。物体検知では、画像中で近くの大きな物体と遠くの小さな物体を共に検知することが望まれる。そこで、バックボーンにおける複数層の特徴量を足し合わせ、様々なスケールの情報を持った新たな特徴量を生成する。この足し合わせを行う部分はネックと呼ばれる。最後にネックで生成された特徴量から物体の存在確率と、物体の大きさを示す枠を予測する。このとき入力画像を区切るグリッドごとに、物体存在確率とその物体のカテゴリおよび物体の枠の候補が求まるが、その中で最も物体の存在確率の高いグリッドを中心とした枠が最終的な検知結果として出力される。この出力部分はヘッドと呼ばれる。本研究ではヘッドにおける物体候補領域の存在確率とカテゴリの情報を用いる。

2.3 点群データの前処理

LiDAR では、レーザーは一定の開口角の範囲内で放射状に照射されるため、レーザーが物体にあたる範囲は、LiDAR から距離が離れるほど広がる。したがって、LiDAR から得られた点群データには、検知対象となる線路周辺の物体以外にも、線路領域外の物体や地面からのレーザーの跳ね返りの分も含まれている。これらの情報は、対象の物体を検知する上で誤検知の原因となりうる。そこで、点群データに前処理を施してこれらの点を除去する。最初に、LiDAR を上下左右の中心として LiDAR の前方(レーザーの照射方向)に直方体の検知領域を設定し、検知領域内の点群データを切り出す。次に切り出した点群データから地面の除去を行う。検知領域をさらに一定サイズの小領域に分割し、その中の点群が持つ高さの情報を参照する。高さの最大値と最小値の差が一定の閾値を下回る小領域内の点群データを除去することにより、高さを持つ物体の周辺から得られる点群のみ抽出できる。

2.4 画像上への点群の投影

点群データと画像の情報を統合するためには、カメラと LiDAR のキャリブレーションを行い、点群を画像上に投影する必要がある。本研究では、キャリブレーションによって求められたカメラと LiDAR の位置関係(外部パラメータ)に基づき点群データを画像に投影する。画像と点群座標の対応関係を図3に示す。画像のある1点の座標を (u, v) 、対応する点群の座標を (X, Y, Z) として両者の関係は

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \dots\dots(1)$$

と表せる。ここで右辺1番目の行列要素はカメラの内部パラメータで、 f は焦点距離、 (c_x, c_y) は画像中心座標を表す。2番目の行列要素はカメラと LiDAR の位置関係を示す外部パラメータで、 $r_{11}, r_{12}, \dots, r_{33}$ は回転、 t_1, t_2, t_3 は並進を表す。本研究では、事前にカメラの内部パラメータを求めた上で、Perspective-n-Point (PnP) 問題を解いて外部パラメータを決定する。PnP 問題では、事前に与えられた画像座標 \mathbf{u} 、点群座標 \mathbf{X} の組から、内部パラメータ行列を \mathbf{P} 、回転行列を \mathbf{R} 、並進ベクトルを \mathbf{t} として、目的関数

$$f(\mathbf{R}, \mathbf{t}) = \|\mathbf{u} - \mathbf{P}(\mathbf{R}\mathbf{X} + \mathbf{t})\| \dots\dots(2)$$

を最小化する。実際には i 番目の画像座標を \mathbf{u}_i 、対応する点群座標を \mathbf{X}_i 、実世界での点群座標の奥行を決める変数を λ_i として、 i 個の方程式

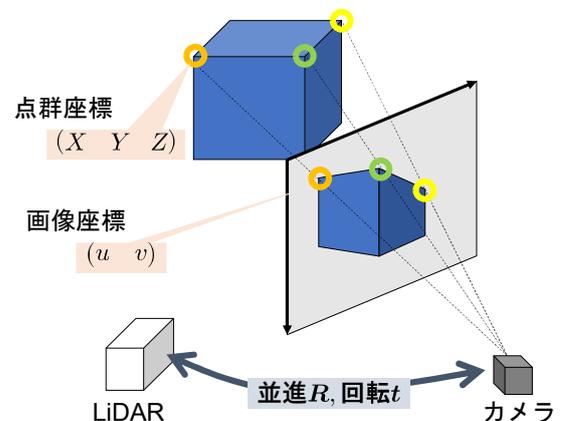


図3 画像座標と点群座標の位置関係

$$\lambda_i \mathbf{u}_i = \mathbf{R} \mathbf{X}_i + \mathbf{t} \dots \dots \dots (3)$$

を解く。画像と点群座標の組の数を N とすると方程式の数は $3N$ 、未知のパラメータ数は $11 + N$ となる。したがって、最低6組の画像と点群座標が与えられていれば、(3)を解くことが可能となる。上記の手続きで求められた外部パラメータを用いて、点群を画像へ投影する。

2.5 画像と点群を統合した物体検知

2.2 節の処理で得られた、グリッドごとの物体の存在確率、カテゴリの情報と、2.4 節の処理で画像に投影された点群の情報を用いて、物体検知判定を行う方法を図4に示す。画像上の物体存在確率を p 、物体カテゴリを c とする。 p はグリッド内の物体らしさ (objectness) p_{obj} 、その物体がカテゴリ c に属する確率を p_c として、

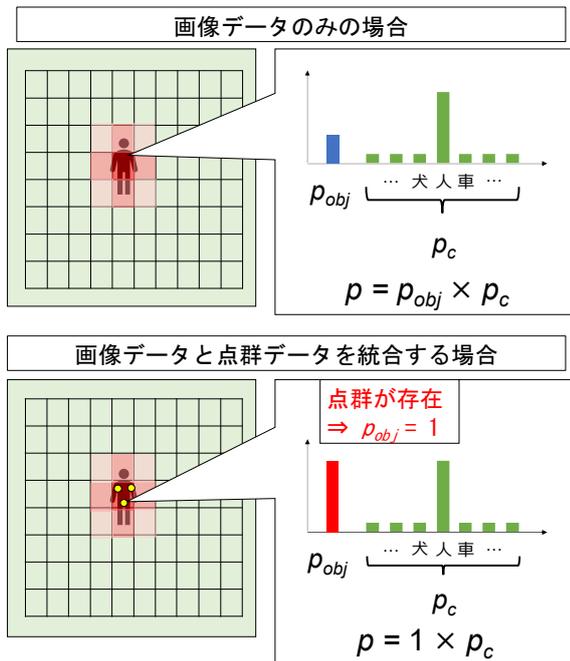


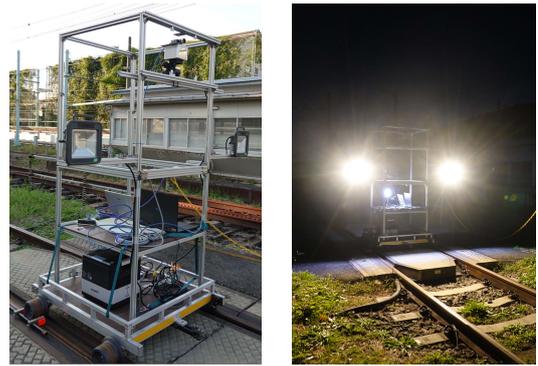
図4 画像と点群を統合した物体検知のイメージ

表1 カメラの諸元

解像度幅[px]	4,096
解像度高さ[px]	2,160
水平画角[°]	31.6
垂直画角[°]	17.0
200m先の空間分解能[mm]	27.6
フレームレート[fps]	10

表2 LiDARの諸元

最大検知距離[m]	500
レーザー開口角(水平)[°]	14.5
レーザー開口角(垂直)[°]	16.2
200m先の空間分解能[mm]	52.3



(a) 昼間 (b) 夜間

図5 測定装置



図6 測定風景

$$p = p_{obj} \times p_c \dots \dots \dots (4)$$

のように算出される。 p_c はそのグリッドにおける p_{obj} が一定の閾値を上回るときに値を持つ。夜間など画像中で物体の特徴が不明瞭な場合、 p_c の値は大きく変化しないが、 p_{obj} の値は大幅に低下するので、最終的な p の値は、主に p_{obj} に依存する。本手法ではグリッドごとに、そのグリッド内で点群データが存在する場合は、確実に物体があるものとみなして p_{obj} の値を1に更新する。この処理によって、夜間でも物体を高い確度で検知することができる。

3. カメラとLiDARの統合手法の精度検証

3.1 データの取得

本手法の精度を検証するために、鉄道総研内の試験線(直線距離約200m)にて、カメラとLiDARを用いた測定を実施した。カメラおよびLiDARのスペックを表1,2に示す。測定にあたって、両センサおよび列車の前照灯を模擬したLEDライトを取り付けた枠を、線路上を走行可能な台車に搭載した装置を用意した。装置の全体像を図5(a)(b)に示す。左右ライトの離れと高さは実際の車両における前照灯の位置関係に合わせた。また、センサの設置高さはおおむね運転士の目線の高さと同しくなるよう調整した。また、カメラとLiDARは両者の光軸が並行になるよう上下に並べて、レールの左右中心の直上に設置した。

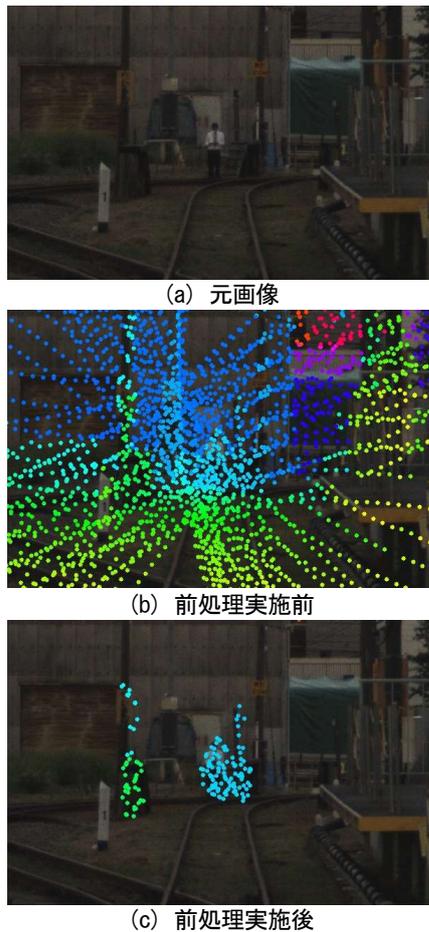


図7 画像に投影された点群データ

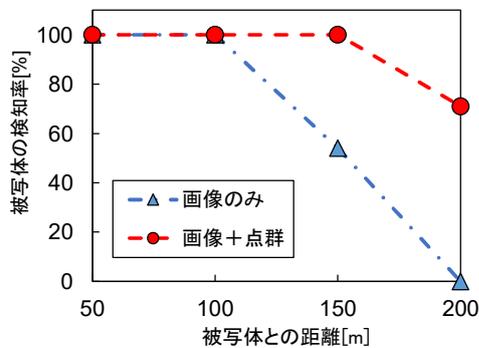


図8 センサ・被写体間の距離と被写体検知率の関係

測定風景を図6に示す。測定時は、測定装置の約200m先に立った被写体（人物）から50m離れた位置で10秒間台車を静止させてデータを取得し、その後50mずつ台車を移動させながら、200m離れた位置まで同様の手続きを繰り返した。また、データ取得時に各測定地点で被写体前方の照度測定を実施した。

3.2 点群データ処理時の条件

取得した点群データから、2.3節に示す方法で地面除去処理を行った。点群データの検知領域は、LiDARを上下左右の中心として幅3.8m、高さ4.3m、奥行は最大205mとした。検知領域の幅と高さは、建築限界内に収まる物体の

みの検知を想定して決定した。また奥行は、試験線の直線区間の終端までが検知領域内に収まるように決定した。地面除去時の点群の高さ方向の分布の閾値は、主に小学生以上の人物を検知することを想定して、1.2mとした。線路内に立つ人物を対象に測定した結果について、前処理（検知領域設定および地面除去）前後の点群データを、2.4節に示す方法で画像上に投影した結果を図7(b)(c)にそれぞれ示す。点群データの色は、被写体までの距離に対応する。前処理により被写体奥の建造物の壁や、人物手前の軌道面に相当する点群が除去され、レールの近傍かつ高さを持つ物体の周辺にある点群のみが抽出されることが確認された。

3.3 人物検知性能の評価結果

2.5節の方法に従って画像データと点群データの情報を統合し、日没直後の時間帯における線路内の人物の検知率を算出した。データ取得を行った時間帯の照度は約1.7lxであった。各地点における検知率は、撮影を行った約10秒間で得られた全フレーム（およそ100フレーム）の画像中で、人物の存在確率 p の値が80%を超えたフレームの割合として定義した。センサと被写体間の距離ごとに、画像データのみを用いて得られた検知率、および画像と点群データの情報を統合して得られた検知率を比較したものを図8に示す。画像のみを用いた場合は、150m先での検知率が54%、200m先での検知率は0%まで低下している。一方、画像と点群データを統合した場合は、150m先での検知率は100%、200m先での検知率は71%となり、提案手法の有効性が確認された。

4. まとめ

線路内の前方支障物検知手法について、主に夜間の検知性能の向上を目指して、カメラの画像データとLiDARの点群データを統合した検知手法を検討した。設定した検知領域内で地面除去処理を行った点群データを画像上に投影し、画像からの検知アルゴリズムで得られた物体候補領域における存在確率の情報と統合することにより、低照度の条件下において、カメラでは検知できなかった200m先の人物が約70%の検知率で検知できるようになることが確認された。今後は点群データから検知領域を自動で設定する方法の検討や、300m以上先での物体に対する検知性能の評価など、本手法の精度向上に向けた取り組みを実施する予定である。最終的には、車上からリアルタイムで検知可能な手法の開発を目指す。

参考文献

- 1) 影山椋, 長峯望, 向嶋宏記, 森下力, 酒井信弘, 福井省三:「深層学習による列車前方カメラ映像からの支障物検知手法の精度評価」, 電気学会 自動車・交通・電気鉄道 合同研究会, TER-20-069 (2020).
- 2) A. Bochkovskiy, C. Y. Wang, H. Y. Mark Liao: "YOLOv4: Optimal Speed and Accuracy of Object Detection", arxiv:2004.10934 (2020).