

降雪環境におけるサーモカメラを用いたリアルタイム人検出システムの開発

江丸貴紀^{*1}, 井内悠介^{*2}, SIRINAKSOMBOON Kittikhun^{*2},
平山英樹^{*3}, 伊藤雅人^{*3}, 西田佳緒理^{*3}, 辻翔太郎^{*3}

1. 緒言

近年、少子高齢化や都市部への人口流出に伴い、地方自治体では労働力不足が深刻な問題となっている。毎年大量の雪が降る札幌市では交通インフラ維持のために除雪作業は必須の作業であるが、この問題の影響を大きく受けている。車両用の道路の除雪作業であれば、主な作業内容は除雪車の運転であるため肉体的な負担はそれほど大きくない。しかし歩道の除雪作業は除雪車の運転の他に、車両の前方を歩いて歩行者や雪に埋もれた障害物の有無等を確認する誘導員が必要となる。この誘導員は深夜に雪の深い道を 10km 以上歩く必要があるため肉体的負担が大きく、担い手不足という問題が起りやすい。このような背景から、歩道除雪作業における誘導員の作業を機械により代替することが求められている。誘導員の役割には、除雪車の周囲にいる歩行者や障害物等を発見しその位置を運転手（除雪機のオペレータ）に知らせることや、歩行者に対し道を開けるよう指示することなどがある。この中でも歩行者の検出というタスクは、RGB カメラと深層学習を組み合わせた様々な手法が提案されており、既に非常に高い精度で実現されている [1]。しかし RGB カメラは雪や雨、霧といったノイズのある悪天候下や、暗闇や逆光などの下では検出精度が低下するという課題がある。除雪作業は夜間に行われ、さらに激しい降雪や地吹雪等も想定されることから、RGB カメラでロバストな人検出を行うことは困難であると考えられる。これに対し、夜間でも周囲環境を認識できるセンサーとしてサーモカメラがある。これは物体自身が放出する赤外線を検出するため RGB カメラのように光源を必要とせず、暗闇や逆光に強いという性質がある(図 1)。サーモカメラを用いることで夜間でも高精度に人物検出を行えることは既に分かっているが [2][3]、激しい降雪のある環境で行われている例はない。赤外線は可視光に比べて波長が長く、霧や煙の中でも散乱しにくいという性質があることから、雪の中も透過してサーモカメラにより検知できると思われる。よって本研究では、歩道除雪作業における誘導員の作業の機械による代替を目的とし、サーモカメラを用いた歩行者検出システムの開発と、サーモカメラの降雪ノイズに対するロバスト性の検証を行った。歩行者検出システムは、サーモカメラと 3D-LiDAR を組み合わせること



(a) RGB camera

(b) Thermal camera

Fig. 1: Comparison of RGB and thermal camera images taken under dark and backlit environments

で検出した歩行者までの距離推定も行い、除雪車の一定範囲内に歩行者が侵入した場合にアラートを発出する。実環境での使用を想定しているため、除雪車に搭載可能な小型コンピュータでリアルタイムに動作するシステムの開発を目指した。また降雪ノイズに対するロバスト性の検証は、歩道除雪環境を模したテストコースを作成して降雪時と非降雪時に実験を行い、それぞれの歩行者検出精度により比較を行った。

2. システム構成

システムの概要を図2に示す。まず初めに、サーモカメラで得られた情報に対して深層学習を用いた物体検出手法を適用することによって歩行者の検出を行う。次に3D-LiDAR から得られる距離情報をサーモカメラの画像平面上に投影し、2つのセンサーの画素の位置関係を対応付ける。そして検出した歩行者までの距離推定を行い、除雪車の一定範囲内に歩行者が入った場合にアラートを発出するというシステム構成になっている。

2. 1 物体検出

歩行者の検出には、深層学習を利用した物体検出手法を用いる。物体検出とは画像内に含まれる物体の位置と種類を特定する手法である。特定した物体 1 つ 1 つに対し、バウンディングボックスと呼ばれる矩形と、物体の種類、推論に対する信頼度を出力する。推論結果の例を図3に示す。深層学習を利用した物体検出は主に Backbone、Neck、Head の 3 つのステップで行われる。一つ目のBackboneは画像から特徴を抽出する役割を持つ。畳み込みフィルタを用いた画像処理を行うと、フィルタに設定されたパラメータの値に応じて画像内のエッジや

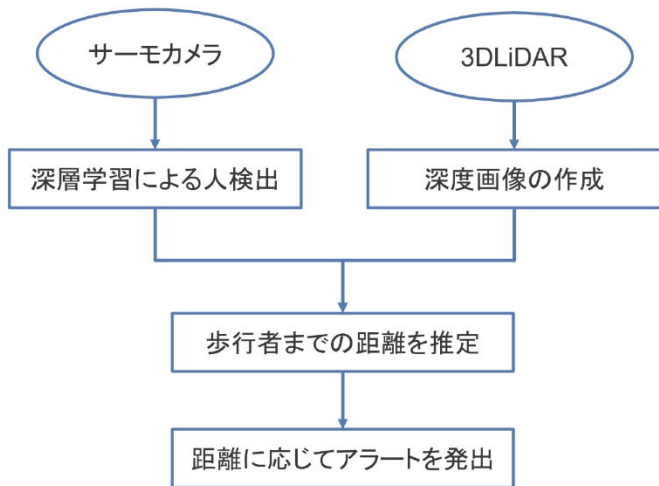


Fig. 2: System overview



Fig. 3: Example of object detection

濃度勾配といった様々な特徴を取得することが出来る。このフィルタをいくつも組み合わせたものが畳み込みニューラルネットワーク (CNN) と呼ばれる構造で、VGGNet や ResNet などこれまで様々なモデルが提案されてきた [4][5]。Backbone ではこの CNN を用いて入力画像から特徴マップの抽出を行う。2 つ目の Neck は、Backbone で得られた異なる解像度の特徴マップの情報を組み合わせる役割を持つ。これにより様々なサイズの物体に対応しやすくなる。3 つ目の Head は得られた特徴から物体の位置とその種類を特定する役割を持つ。本研究では物体検出の手法として YOLOv5 [7] を用いる。YOLO は非常に高速に動作する物体検出モデルとして 2015 年に最初のモデルが公開され [6]、それ以降も様々な改良が加えられてきた。本研究で想定している歩道除雪環境は、精度に加えて処理速度も重要視されるためこのモデルを採用した。YOLOv5 の構造は、Backbone に CSPDarknet、Neck に SPP (Spatial Pyramid Pooling) と PAN (Path Aggregation Network)、Head に YOLOv3 が用いられている。また、YOLOv5 には s, m, l, x の複数のバージョンがあり、s は計算量が少なく精度は低い、x は計算量が多く精度は高いといった特徴がある。

2. 2 センサフュージョン

本研究ではサーモカメラと 3D-LiDAR のセンサフュージョンを行うことで、検出した歩行者までの距離情報を取得できるようにした。(1)式により、LiDAR から得られる点群を画像平面へ投影することができる。(1)式の左辺 (u, v) は画像平面の二次元座標を表し、式の右辺 (X, Y, Z) はワールド座標系での三次元座標を表している。右辺第一項はカメラの内部パラメータを表し、 (c_x, c_y) は画像の中心座標、 (f_x, f_y) はそれぞれカメラの水平・垂直視野角から得られる焦点距離である。右辺第二項は同次変換行列で、3D-LiDAR に対するサーモカメラの相対位置と回転を

表す。サーモ画像平面上に投影する際に、点群情報は距離情報に変換する。LiDAR から得られる距離値を 8bit の値に変換し、LiDAR に近い点を黒、LiDAR から遠い、もしくは点群のない範囲を白として深度画像を作成すると、図4右図のような画像が得られる。LiDAR の垂直視野角がサーモカメラの垂直視野角よりも狭いため、点群情報は画像の縦方向中心付近に寄っている。左のサーモ画像と右の深度画像は位置関係が対応しているため、サーモ画像内の物体までの距離を求めることができる。

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

2. 3 アラート発出

開発するアラートシステムは、歩行者が除雪車の一定範囲内に侵入した際にアラートを発出することを想定している。ここではアラートを発出するかどうかの基準となる距離の算出方法について説明する。一般に車が止まるまでに必要な停止距離は、ブレーキを踏み始めるまでに進む空走距離と、ブレーキを踏んでから停止するまでに進む制動距離の和により求められる。空走距離は反応時間 (秒) × 車の速度 (m/s)、制動距離は車の速度 (km/h) の 2 乗 ÷ (254 × 摩擦係数) の式で算出される。除雪車の走行速度を 10km/h、反応時間を 2 秒、摩擦係数を一般的な圧雪路で用いられる 0.15 とすると、空走距離は 5.6m、制動距離は 2.6m、停止距離は 8.2m となる。除雪車の運転席前方には、オーガと呼ばれる雪をかき込むスクリー型の機械がついており、運転席からその先端部分までの距離は 2.1m ある。以上よりシステムは、歩行者が運転席から 10.3m 以内に入るまでにアラートを発出する必要がある。余裕をもって停止できるように、アラート発出を行う閾値を 15m とする。

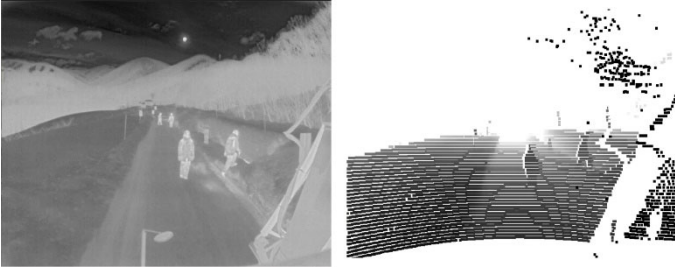


Fig. 4: Example of sensor fusion. The left image shows the data obtained by the thermal camera and the right image shows the data obtained by the LiDAR.



Fig. 5: Experimental equipment installed on the roof of a snowplow

3. 実験

3. 1 実験装置

実験装置の外観を図5に示す。実際に歩道除雪作業で使われている小型ロータリー車の上部、左右 2 か所にサーモカメラ、3D-LiDAR、RGB カメラを取り付けた。排雪時に立てるシューターで視界が遮られてしまうため、視界を確保するために左右に同じものを取り付けている。サーモカメラと 3D-LiDAR の仕様はそれぞれTable 1、Table 2 の通りである。

Table 1: Thermal camera specification (FLIR ADK) [8]

Resolution	640 × 512 px
Horizontal view angle	75°
Frame rate	30Hz
Spectral region	8~14μm

Table 2: LiDAR specification (Ouster OS1-64) [9]

Vertical resolution	64 channels
Measurement distance	120m
Vertical view angle	45° (±22. 5°)
Frame rate	10Hz



Fig. 6: The test course at NICHIO Corporation, where the experiment was conducted.

3. 2 実験内容

実際の歩道除雪環境では、歩道の両側に 1~2m ほどの雪壁ができる。そのため雪壁によって歩行者の体が隠れてしまい、積雪のない環境と比べて歩行者検出の難易度は高くなると考えられる。安全に配慮した上で実験データを効率的に取得するため、その上で実環境に近い環境で評価を行うため、雪壁や看板、スノーポール等を設置したテストコースを図6のように作成した。実験は、センサー類を取り付けた除雪車の前方5mから40mの範囲で人が歩く様子を撮影するという方法で行った。また、立位状態の人間だけではなく、臥位状態の人間についてもデータ取得を行った。この実験を、除雪車の位置や向き、走行状態・停止状態、雪を排出するシューターの有無、天候、時間帯等の条件を変化させながら複数回行い、データを取得した。

4. 結果

4. 1 降雪環境と非降雪環境の比較

実験時の様子を図7に示す。降雪のある環境では強風が吹いており、路面の雪も吹き上げられていたためかなり視界が悪い状況だった。除雪車に搭載した RGB カメラとサーモカメラで撮影した画像をそれぞれ図8および図9に示す。RGB画像の場合、降雪が多い環境では車両からの光が雪に反射して見えにくくなっており、テストコース上の人物を視認するのは困難である。それに対しサーモカメラでは雪があまり映り込んでいないため、人物を認識することが可能である。しかし、これよりさらに降雪の多い時や、除雪車のシュートから排出される雪に対しては視界が遮られてしまった。今回の実験から、赤外線は雪による散乱が起こりにくいため、サーモカメラの方が降



Fig. 7: Experiments during snowfall and at night

雪環境でも周囲環境の認識はしやすいが、密度の高い雪で完全に光が遮られた場合、サーモカメラでも視認性が低下すると言える。

4. 2 データセットの作成

深層学習のモデルの学習と評価を行うために、これらの状況で取得されたデータからデータセットを作成した。元のデータは動画として取得されているため、これらからフレームを抜き出し、合計1000枚の画像を抽出した。このデータセットを用いて歩行者の検出を行うため、画像は人が写っているものに限定し、さらに時間帯（昼間・夜間）や天候（晴れ・降雪）が均等になるように画像を選択した。また、実際の環境では歩行者が除雪車の運転席から見やすい状況にあるとは限らない。そのため、どのような状況であっても人を検出できるようにモデルのトレーニングを行う必要がある。このような理由から、身体の一部が隠れている状態や、座っていたり寝ていたりといった普段と異なる体制のデータを優先的にデータセットに取り入れた。

データセットの作成のためにはアノテーション[10]と呼ばれる作業が必要である。アノテーションとは各画像に含まれる物体をバウンディングボックスで囲み、各ボックスに対し人や車と行ったラベル付けを行う作業を指す。本研究では、アノテーションを行うためにLabelImg[11]というツールを利用した。このようにアノテーションを行ったサーモ画像を1000枚準備し、それらを学習用800枚、テスト用200枚に分割した。このうちテスト用の画像は、降雪があるものとないもので分けてそれぞれ歩行者検出を行い、その結果を定量的に評価した。

4. 3 モデルのトレーニング

作成したデータセットを利用してモデルの学習を行っ



Fig. 8: RGB camera information in heavy snowfall environments

た。深層学習では損失関数を設定し、それが最小の値をとるように学習が行われる。試行錯誤の結果、epoch数を100、batch sizeを16として学習を行った。学習にはおよそ1時間ほどかかった。なお、学習に使用したコンピュータのスペックは、以下の通りである。OS: Windows 10、CPU: AMD Ryzen 5 3600 (6 Core Processor)、GPU: GeForce RTX 2070 Super (memory: 8GB)、memory: 32GB。

4. 4 テスト結果

次に降雪環境と非降雪環境それぞれで取得したデータに対し、歩行者の検出を行った結果を Table 4 に示す。評価指標には AP (Average Precision) を用いており、この値が大きいほどより正確に検出できていると言える。AP50 については降雪時と非降雪時で差はなく、AP50:95 では 5%程度降雪時のほうが低いという結果が得られた。この結果から降雪時には推論したバウンディングボックスの位置精度が多少落ちるものの、サーモカメラが降雪ノイズに対してもロバストに機能することが確認できた。

さらにYOLOv5の各バージョンの評価であるが、最も小さいモデルであるYOLOv5sであっても他のバージョンとの性能差はさほど大きくなく、処理時間を見ると最も短い時間で検出が実現できている。この結果から、本研究ではアラートシステムの歩行者検出手法としてYOLOv5sを採用する。

Table 3: Jetson AGX Xavier specifications[11]

GPU	512-core Volta GPU
CPU	8-core ARM v8.2 64bit CPU
Memory	32GB 256bit LPDDR4x
Size	105 × 105 × 65 mm

Table 4: Comparison of accuracy on clear and snowy days with a quantitative evaluation using AP

Model	Clear day		Snowy day	
	AP	AP50	AP	AP50
YOLOv5s	74.4	99.1	71.4	99.5
YOLOv5m	74.5	99.4	71.7	99.5
YOLOv5l	74.7	99.3	72.7	99.5
YOLOv5x	74.2	99.3	71.9	99.5

Table 5: Differences in computation time among the four models of yolov5

Model	Processing time [ms]
YOLOv5s	35
YOLOv5m	65
YOLOv5l	110
YOLOv5x	190

4. 5 処理時間

YOLOv5 の各モデルで推論にかかった時間をTable 5に示す。処理はNVIDIA 社が提供する小型コンピュータ (Jetson AGX Xavier Developer Kit) 上で行った。Jetson AGX Xavierの仕様をTable 3に示す。歩行者の検出以外の主な処理では、深度画像の作成に1ms未満、歩行者までの距離推定に3msほどかかった。歩行者の検出にYOLOv5s を採用した場合、処理全体としてかかる時間が40ms程度であるため、25fps (frame per second、1秒当たりの処理回数) 程度のフレームレートで動作させることができる。LiDARのフレームレートが10fps、サーモカメラのフレームレートが30fpsであることから、実用化する上で十分な処理速度である。

本アラートシステムでは、除雪車の視界を確保するために図5に示すように除雪車の上部にセンサーを二組搭載しており、それぞれに対して一連の処理を行っている。左右のセンサーの情報を併せて処理を行った場合の処理速度は9fpsとなった。処理時間が2倍よりも長くなった原因としては、2組のセンサーを用いることで新たな処理が増えたことや、処理するデータ量の増加によるコンピュータの処理効率の悪化などが考えられる。しかしながら、除雪機の移動速度を勘案すると2fps程度の処理速度が実現できていれば十分実用的であると考えられる。さらにこの処理速度はLiDARのフレームレートと同等の速度であるため、本研究で開発したアラートシステムは、実環境での使用を想定しても十分なリアルタイム性を持っていると言える。



Fig. 9: Thermal camera information was obtained simultaneously with the data in Fig. 8 and applied to our proposed method



Fig. 10: Example of alert system output

4. 6 アラートシステム

サーモカメラと3D-LiDAR のセンサフュージョンを行うことで、検出した歩行者までの距離を精度よく求めることが出来た。本研究で開発したシステムでは、推定した距離に応じて検出した歩行者を囲うバウンディングボックスの色を変更し、一定範囲内に歩行者が侵入した際に文字を表示させることでアラートを発出する。出力例を写真に示す。除雪車から遠い時は緑、一定範囲内に入ったときは橙、さらに危険な範囲に入ってきた場合は赤色で表示している。本研究では出力画像に表示させるだけにとどまっているが、実用化する上ではスピーカーや緊急停止装置といった外部機器と接続することで、より効果的なアラートシステムになると考えられる。

5. 結言

本研究では歩道除雪環境で用いる歩行者検出アラートシステムの開発と、降雪ノイズに対するサーモカメラのロバスト性の検証を行った。サーモカメラと3DLiDARを組み合わせて、高精度に前方にいる歩行者までの距離を取得することが出来た。物体検出手法にYOLOv5sを採用することで、Jetsonのような小型コンピュータでもリアルタイムで処理を行うことができた。また実験結果から、サーモカメラは降雪ノイズの影響を受けずに歩行者の検出を行えることが分かり、視界不良が想定される歩道除雪環境でも安定した動作が可能であると言える。現状の課題として、誤検出や距離推定の誤りがあった際もアラートが出されるため、アラートの信頼性が十分ではないという問題がある。今後はアラート発出までのアルゴリズムを改良し、アラートの信頼性をより高めることに取り組んでいく。

謝辞

本研究の一部は、JSPS科研費24K07413、および株式会社NICHIJ0との共同研究「歩道除雪機用人検出アラートシステムに関する研究」により実施されたものである。また、一部の実験データは株式会社NICHIJ0テストコースによって取得されたものであり、その取得に際して多大なるご協力を頂いた。ここに記して感謝の意を表す。

参考文献

- [1] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, Stan Z. Li, Occlusion-aware R-CNN: Detecting Pedestrians in a Crowd, in Proc. the European Conference on Computer Vision (ECCV), pp. 637-653, DOI: 10.1007/978-3-030-01219-9_39, 2018.
- [2] Jeonghyun Baek, Sungjun Hong, Jisu Kim, and Euntai Kim, Efficient Pedestrian Detection at Nighttime Using a Thermal Camera, Sensors, 17(8), DOI: 10.3390/s17081850, 2017.
- [3] Eun Som Jeon, Jong-Suk Choi, Ji Hoon Lee, Kwang Yong Shin, Yeong Gon Kim, Toan Thanh Le and Kang Ryoung Park, Human Detection Based on the Generation of a Background Image by Using a Far-Infrared Light Camera, Sensors, 15(3), pp. 6763-6788, DOI: 10.3390/s150306763, 2015.
- [4] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in Proc. 3rd International Conference on Learning Representations (ICLR 2015), pp. 1-14, DOI: 10.48550/arXiv.1409.1556, 2014.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, in Proc. Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2016.90, 2015.
- [6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, in Proc Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2016.91, 2015.
- [7] GitHub - ultralytics/yolov5, <https://github.com/ultralytics/yolov5> (accessed 2024-9-26).
- [8] FLIR ADK 赤外線ビジョン自動車用開発キット. Teledyne FLIR, <https://www.flir.jp/products/adk/> (accessed 2024-9-26).
- [9] 高解像度OS1 LIDAR センサー. Ouster, <https://ouster.com/jp/products/scanning-lidar/os1-sensor/> (accessed 2024-9-26).
- [10] Eleanor Watson, Thiago Viana, and Shujun Zhang, Augmented Behavioral Annotation Tools, with Application to Multimodal Datasets and Models: A Systematic Review, AI, 4(1), DOI: 10.3390/ai4010007, 2023.
- [11] heartexlabs/labelImg, <https://github.com/HeartEx-labs/labelImg> (accessed 2024-9-26).
- [12] Jetson AGX Xavier Developer Kit. NVIDIA Developer, <https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/jetson-agx-xavier/> (accessed 2024-9-26)