

# LiDAR とカメラのセンサーフュージョンによる 点群からのノイズ除去

板倉 健太<sup>1</sup>・林 拓哉<sup>1</sup>・上脇 優人<sup>1</sup>・全 邦釘<sup>2</sup>

<sup>1</sup> ImVisionLabs 株式会社 (〒113-8485 東京都文京区本郷 7-3-1)

東京大学南研究棟アントレプレナーラボ)

<sup>2</sup> 東京大学大学院工学系研究科総合研究機構

E-mail: chun@g.ecc.u-tokyo.ac.jp (Corresponding Author)

近年、様々な分野で 3 次元点群の利活用が進んでいる。3 次元点群を得るための手法として、地上型レーザーキャナを利用する方法がある。寸法に関する情報は精度が高いというメリットがある一方で、地上にて計測を行うため、人や自動車などの移動体のノイズが計測されることが多い。地上型レーザーキャナは、点群と同時に画像も取得されることが多い。そこで、本研究では地上型レーザーキャナを用いた点群計測とカメラ画像とのセンサーフュージョンを通じて人のセグメンテーションを行う。SOLOv2 を利用して、画像から人のセグメンテーションを行った。そして、その情報をカメラの内部パラメータや外部パラメータを利用して、点群にマッピングする。これにより、精度良く人のノイズを点群から除去することができた。

**Key Words:** Cross calibration, LiDAR, Noise removal, Point cloud, Sensor fusion

## 1. はじめに

近年、3 次元点群の利活用が様々な分野で進んでいる。3 次元点群とは、xyz座標を持つ点の集まりによって、対象の 3 次元構造を表現するデータ形式である。山下ら<sup>1)</sup>は、Mobile Mapping System (MMS) で取得した点群を使用して、建築物の測量や、コンクリートの浮きや剥離を検出するシステムを開発した。類似した例として、大伴ら<sup>2)</sup>の MMS を使用した、損傷検出があげられる。この研究では、事前に点群上で、作業に使用する重機のシミュレーションを行うことで、作業時間短縮や、効率化が見込まれると報告されている。その他の事例として、送電線への利用が挙げられる。Chen ら<sup>3)</sup>は、航空レーザーキャナで取得した点群から、送電線および森林中の植生を検出した。そして、両者の距離を算出することで、送電線に接近した樹木を素早く特定する手法を開発した。これによって、従来は目視で確認されていた作業を自動で行うことができるようになる。

また、考古学においても 3 次元点群が広く用いられている。Inomata<sup>4)</sup>らは、航空レーザーキャナを用いて、広範囲にわたるメソアメリカ文明の遺跡を識別した。この事例は、LiDAR (Light Detection and Ranging) と呼ばれる

レーザーキャナを使用することで、遺跡を効率的に見てできることを示した。このような人工構造物を対象とした事例のほかにも、自然植生といった対象を例に 3 次元点群を活用した例も多い。例えば、地上にて LiDAR を利用して取得した 3 次元点群から、樹木を自動的に検出する方法が報告されている。これにより、胸高直径や樹高などの情報も自動的に取得できることが示唆され、植生の計測においても 3 次元点群の利用が有用であることが示された<sup>5,6)</sup>。また、深層学習を利用した手法や点群処理を利用し、都市部の街路樹を自動的に精度良く検出できることも報告されている<sup>7)</sup>。

3 次元点群を取得する方法としては、LiDAR を利用するものや、Structure from Motion (SfM) と Multi-View Stereo (MVS) 技術を用いた、SfM-MVS を利用するものなどがある。LiDAR はレーザーを用いて環境をスキャンするため、木の葉や枝を部分的に透過し、LiDAR から見て下方にある地表面や奥側にある物体の情報を捉えることが可能である。また、LiDAR では、レーザーを対象に反射し、返ってくるまでの時間を計測することで、対象までの実距離および、3 次元構造を把握する。そのため、画像から対象の 3 次元構造を把握する、SfM-MVS と比べて、寸法に関する情報は精度が高いといったメリットもある。

LiDARは、その応用範囲の広さから、様々なタイプのプラットフォームを用いて計測が行われている。また、計測対象やプラットフォームによって、使用されるLiDARの種類も変化する。例えば、航空機や UAV (Unmanned Aerial Vehicle) に搭載し、上空から計測するものがある。このタイプの LiDAR は広範囲を計測する能力を有するが、計測精度には限界があり、またコストが高い傾向にある。一方で、地上型レーザースキャナ (TLS: Terrestrial Laser Scanners) は、対象の3次元形状を高精度で取得できるというメリットがある。特に細かな形状把握を必要とする測量や 3D モデリングの分野で有用である。

しかし、地上型レーザースキャナで取得したデータを処理する際には、ノイズの存在が大きな課題となる。地上でのデータ取得は、建造物の計測を目的としている場合でも自動車や人などの移動する物体が点群データに含まれることが多い。これらの計測対象でない物体は、計測データの品質に悪影響を及ぼすため、点群データからこれらのノイズを除去する作業が必要となる。従来は、手動でノイズ除去処理が行われてきた。このノイズ除去プロセスは、時間と労力を要する作業であり、効率化のための技術開発が求められている。

自動的にノイズを除去する方法としては、以下のようなものがある<sup>9)</sup>。例えば、各点に対して、近傍点を  $n$  点取得し、それぞれの近傍点までの距離を求める。その計算を全ての点に対して行い、近傍点までの距離が比較的遠い点をノイズとする方法である。これは、ノイズの点が周辺から孤立した形で存在しているという仮定に基づき、ユーザーが設定した閾値よりも近傍までの距離が遠い点をノイズとみなすものである。また、その他の手法の例として、一定の探索範囲に従い、点群を複数のクラスタに分けて、そのクラスタを構成する点数をカウントする。そして、そのクラスタを構成する点数が少ない場合は、小さなノイズの塊とみなし、除去するといった手法も存在する。これは、ノイズの点は小さなクラスタとして存在しており、一方で、建造物などは密度が高く、点数も多いということに基づいている。ノイズ除去の方法として、深層学習を利用した方法も知られている。PointCleanNet<sup>9)</sup> は、点群から、物体形状を抽出することで、物体形状にかかわりがない点、つまりノイズを除去する手法である。しかし、これらの手法では除去しきれない対象も存在する。例えば、人や自動車といった移動体のノイズは、ある程度の点群密度を有するため、先述した手法では、それらの点を除去することは難しい。また、上記の方法の他に、色情報によってノイズを除去する手法もあるが、対象の色が毎回異なる場合があるため、安定したノイズの識別は困難である。また、高さ情報によってノイズを除去する場合でも、目標物の一部が点群

として形成されず、高さが変動する可能性があるため、安定してノイズを除去することは難しい。このように、自動でノイズを除去する手法は存在するものの、いずれも精度良くノイズを除去することは困難な場合も多い。また、どの手法も、点の密度や、高さなどの点群に付随する情報のみを利用してノイズ除去を行っている。ノイズ除去を行う際に、点群だけでなく、画像情報などの他の情報も利用することで、ノイズ除去の精度が向上する可能性がある。

地上型レーザースキャナでは、点群の色付けなどの目的で、点群計測の際に、内蔵されたカメラにて画像も同時に取得されることが多い。その画像を取得し、点群と統合して解析することで、ノイズの除去も精度良く行える可能性がある。点群と画像のクロスキャリブレーション (Cross Calibration) を行うことで、これら二つの異なるタイプのデータを統合 (センサーフュージョン) し、両方のセンサー情報の強みを生かすことができる。例えば、点群データから、3次元空間における物体の位置や形状に関する正確な情報を得ることができる。一方、画像データからは、色やテクスチャーに関する詳細な情報を得ることができ、その画像中に映る人などの物体の位置を特定することができる。

画像から物体のセグメンテーションを行う深層学習ネットワークは数多く報告されている。例えば、画像のピクセルごとにカテゴリを分けるセマンティックセグメンテーションの手法では、DeepLabv3+<sup>10)</sup> などが有名である。劉ら<sup>11)</sup> は DeepLabv3+ を活用し、腐食損傷を自動的に検出する試みを行った。さらに、ファインチューニングを適用することで、境界線部分での検出精度の向上を報告した。ピクセルごとにカテゴリの分類を行うだけでなく、個体ごとに分離をするインスタンスセグメンテーションの手法も知られている。例えば、Mask R-CNN<sup>12)</sup> や SOLO<sup>13)</sup>、SOLOv2<sup>14)</sup> などが広く利用されている。山根ら<sup>15)</sup> は Mask R-CNN を利用して、3次元モデルの中の損傷の部位のセグメンテーションを行った。また、高解像度画像からメモリ効率良く、画像のセグメンテーションを行う方法なども報告されている<sup>16)</sup>。このように深層学習モデルを利用して画像のセグメンテーションを効果的に行った例が多く報告されている。また、深層学習モデルを特定の課題に利用するためには、課題に適した少量の学習データで再学習させるファインチューニングといった手法が必要である。しかし、人や自動車といった頻出のカテゴリに対して使用する場合は、大規模なデータセットで学習した学習済みモデルをファインチューニングすることなく利用できる可能性がある。

そこで、本研究では、地上型レーザースキャナを利用

し、3次元点群の計測を行う。その際に点群として計測された人を除去する方法を提案する。本研究の実験では、地上型レーザースキャナにより取得したデータから、画像と点群を取得する。事前準備として、チェッカーボードを利用してカメラの内部パラメータや外部パラメータを求めた。そして、それらのパラメータと計測した点群とレーザースキャナの位置関係を考慮し、カメラ画像と点群のセンサーフュージョンを実行した。そして、画像を対象として、深層学習によるインスタンスセグメンテーションの手法を利用して、人のセグメンテーションを行う。これにより、画像にてセグメンテーションした人の情報を点群にマッピングする。これらの処理を通して、点群に含まれるノイズの除去を試みた。

## 2. 実験試料と方法

### (1) 点群の計測について

本研究でのLiDAR計測は、2023年12月に福島県石川郡平田村にて行った。橋梁、河川、および樹木が混在する環境で実施した。周辺には人が存在したため、得られた点群中にも人が記録されている。計測には、Matterport Pro3 (Matterport, Inc., USA) を使用した。また本論文ではこのLiDARをMatterportと呼ぶ。レーザービームの波長は904 nmであり、視野角は水平360度、垂直295度である。0.5 mから100mの範囲の対象を計測することができ、測距精度は10mで±20 mmである。また、1秒当たり10万点取得可能であり、一回のスキャンが20秒未満と比較的短い時間で完了する点の特徴である。点群と同時に、4096×4096 pixelのカラー画像を取得することができる。本体の重量は2.2 kgであり、三脚の上にMatterportを設置し、計測を行った。本LiDARでは、水平方向に関しては、前後左右の4方向、また、鉛直方向には、上下の2方向の画像撮影を行う。なお、本対象地では、上下方向に計測対象がなかったため、解析には、水平方向の画像のみを利用した。

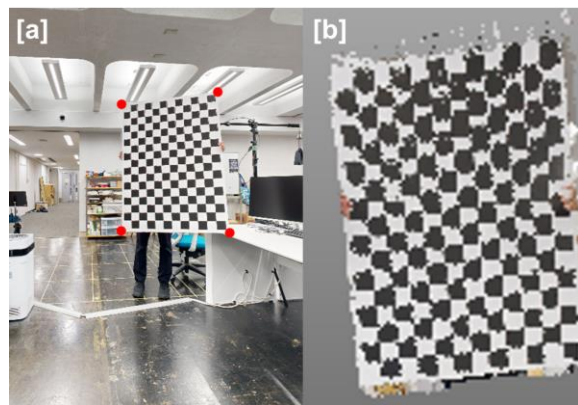


図-2 LiDAR とカメラのクロスキャリブレーションを行っている様子。[a]では、チェッカーボードの角を画像から検出しており、これらの角は赤い丸でマークされている。一方、[b]の部分では、Matterport でエクスポートされたチェッカーボードに対応する点群データを手作業で選択し、別ファイルとして保存している。

取得した点群は、E57形式で出力した。E57形式には画像や点群の階層があり、点群データだけでなく、計測時に取得した画像も取得することができる。さらに、そのスキャンの点群の位置や向きもクォータニオンにて保存されている。本研究では、この位置関係を利用し、LiDARとカメラのデータを統合する。

### (2) LiDAR-Camera キャリブレーションについて

本節では、LiDAR-Cameraのキャリブレーション手法について述べる。なお、本実験で使用した手法は、Jiaoら<sup>17)</sup>により報告されている手法を参考にした。後述する、LiDARとカメラのセンサーフュージョンのために、あらかじめ、カメラの内部パラメータおよび外部パラメータを求めておく必要がある。そのため、本節では、室内にて点群計測およびを行い、キャリブレーションを行った。

#### a) チェッカーボードの撮影

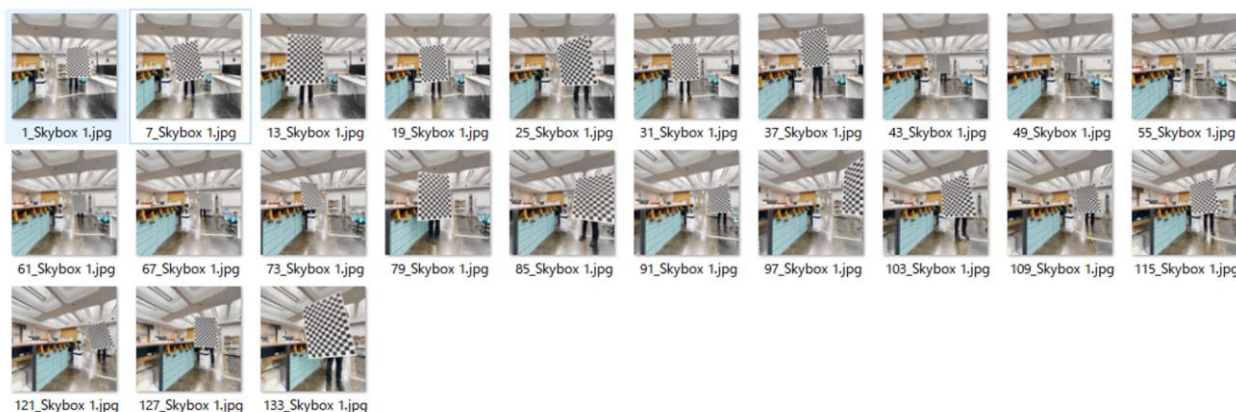


図-1 Matterport に内蔵されているカメラを利用して、チェッカーボードを撮影した時の様子。

Matterportに内蔵されているカメラにてチェッカーボードの撮影を行った。点群を取得しながら画像の撮影を行った。チェッカーボードは、A0のボードに白黒のパターンを印刷し、作成した。チェッカーボードを手で持ち、様々な角度から撮影を行い、図-1のような合計23枚の画像を得た。

これらの画像を利用して、カメラキャリブレーションを行い、内部パラメータの推定を行った。キャリブレーションにおいては、MATLAB 2023bのカメラキャリブレーターを利用した。画像を入力し、設定値を選択することで自動的にカメラキャリブレーションを行うことができる。本研究では、イメージのゆがみ「低」とし、半径方向の歪みの補正には、4次多項式を利用した。上記の処理を、前後左右で取得した画像それぞれに対して、計4回行った。

### b) チェッカーボードの撮影

次に点群と画像情報のセンサーフュージョンを行う。図-2の[a]のように、画像からチェッカーボードの角を検出する。ここでは角が赤丸で示されている。さらに、[b]に示すような、チェッカーボードに対応する点群をMatterportからエクスポートしたファイルから手動にて切り取り、別ファイルとして保存した。本項の処理では、画像上のチェッカーボードと点群のチェッカーボードの角の対応付けを行う。

ここでは、世界座標をカメラの中心を原点としたカメラ座標に変換するために、世界座標の点群のxyz座標を回転・移動させることを考える。例えば、ある点をx軸、y軸、z軸まわりにそれぞれ、 $\alpha$ 、 $\beta$ 、 $\gamma$ 回転させるとき、その処理は、以下のような行列で表すことができる。

$$R = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} \quad (1)$$

同様に、この回転行列Rに加え、並進を行うためには、以下のような4x4の行列を利用する。これにより $X_w, Y_w, Z_w$ で表される世界座標の値をカメラ座標に変換することができる。ここで、 $t_1, t_2, t_3$ はそれぞれx, y, z座標に関する並進のためのパラメータであり、 $r$ は回転行列のパラメータである。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2)$$

本研究では、23組のチェッカーボードの画像とその点群のペアを利用し、式(2)の外部パラメータ行列を求めた。

### c) チェッカーボードの撮影

さらに、以下の式(3)のように、カメラの内部パラメータ行列を左からかけることで、カメラの画像座標に変換することができる。ここで、 $f_x, f_y$ はそれぞれ、x, y方向に関する焦点距離であり、 $c_x, c_y$ はカメラの光学中心を示す。また、 $s$ と $t$ はせん断、およびスケールに関するパラメータである。式(3)を透視投影変換行列と呼び、この行列を利用することで、点群の座標と画像座標の画素値の対応関係を求めることができる。

$$t \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

### (3) 点群のノイズ除去

#### a) 画像のセグメンテーションと点群へのマッピング

カメラ画像とLiDARから取得した点群データのセンサーフュージョンを利用して、点群内の人に関連するノイズの点を効果的に分類する手法を図-3に示す。本手法では、まず初めに2次元画像上でインスタンスセグメンテーションのアルゴリズムであるSOLOv2<sup>14)</sup>を適用し、人の輪郭を抽出する。SOLOv2はシンプルな構造を持ち、高速に推論を行うことができる。3次元点群の処理は時間がかかることが多いため、高速に実行できるアルゴリズムを選択した。ここで得られたセグメンテーション結果は、先述した、カメラとLiDARのクロスキャリブレーションの過程で得られた、カメラの内部および外部パラメータを用いて3次元空間にマッピングされる。これにより、画像上にて識別された人を、LiDARの点群デー

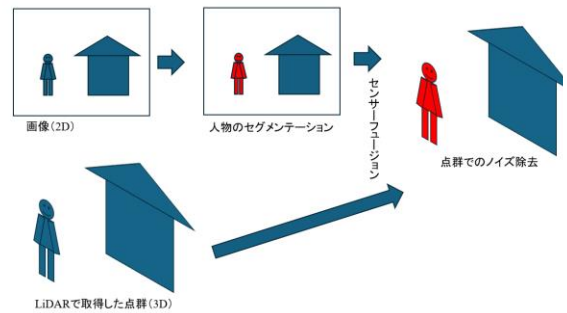


図-3 LiDAR とカメラの情報を統合し、点群でのノイズ分類を行う時の様子。2次元画像に対してインスタンスセグメンテーションを行い人の輪郭を特定する。この手順で得られたセグメンテーションの結果は、以前に行われたカメラとLiDARのクロスキャリブレーションで得られた内部および外部パラメータを使用して、3次元空間へマッピングされる。これにより、画像内で特定された人がLiDARの点群データと結びつけられ、点群データ内でのノイズとみなされた部分は赤色で示される。

タに対応付けることが可能となる。ここでノイズとして分類した点群は赤で着色することで可視化を行う。また、利用した SOLOv2 ネットワーク<sup>14)</sup>は、バックボーンに、ResNet50<sup>18)</sup>を利用しており、訓練は COCO データセット<sup>19)</sup>を利用したものである。この画像のセグメンテーションや後続するノイズ除去の処理は MATLAB 2023b (MathWorks, USA) を利用して行われた。マッピングを行う際は、画像の膨張処理を行い、画像の輪郭およびその領域を膨張させた。人として認識された画像での領域を広げノイズ除去の漏れを防ぎ、さらにスキャンの際に人が移動する場合にも対応するためである。また、セグメンテーションの結果を点群にマッピングした際に複数のノイズのクラスタに分かれることがある。その理由として、マッピングにずれがあると、対象の物体の奥側の領域までノイズの情報が付与されるためである。そこで、ノイズの情報を投影した 3 次元点群のノイズの点を 50 cm の範囲でクラスタリングし、その中で最も LiDAR に近いクラスタをノイズの点とした。人の点群が LiDAR から遠い距離にあっても、その点群の点間距離は 50 cm よりも小さかったためである。

#### b) 画像のセグメンテーションと点群へのマッピング

ノイズ除去の精度評価のために、対象地にて計測された点群データを 3 ファイル用意した。これらの点群データには、それぞれ人が映り込んでいる。次に、画像を参照しながらこれら人の点群を特定し、手動による切り取りを行った。そして、手動により切り取られたファイルと、本手法でノイズとして分類された点群とを比較することにより、精度評価を実施した。この評価は合計 9 つのノイズを用いて行われた。ノイズ分類の精度評価には、Recall, Precision, および F-measure を利用した。精度評価

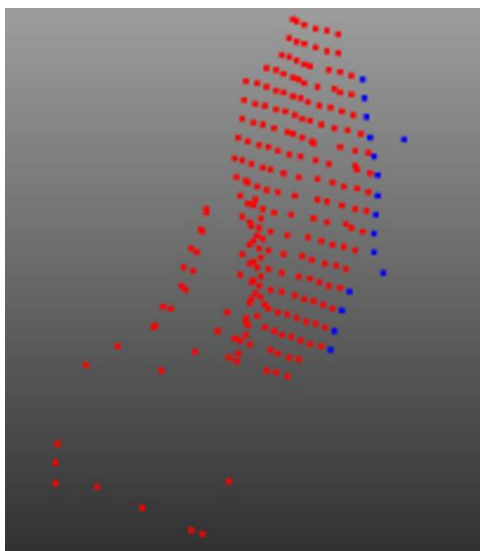


図4 ノイズ除去の例。赤点が本手法によりノイズと分類した点を示している。青点は、ファイルから手動にてノイズの点を切り出したものである。

においては、以下のように、手動にて人のノイズの点を切り出した。図4では、本手法によりノイズと分類した点を赤色、実際にノイズである点を青色で表している。両者は同じ座標系を有している。ノイズと分類した点のそれぞれに対して、正解データにその点が含まれるかを確認し、Precision を計算した。また、正解データのそれぞれの点が、ノイズとして分類されたかどうかを検証し、Recall を計算した。そしてこれらの結果を利用し、F-measure の値を計算した。

### 3. 結果と考察

#### (1) LiDAR-Camera キャリブレーションの結果

##### a) カメラキャリブレーションについて

図5はカメラキャリブレーションをしている様子を示す。チェッカーボード中のパターンが自動的に検出されていることがわかる。この画像以外に関しても、自動的にチェッカーボードの角を検出することができた。しかし、一部のサンプルに関しては、角を検出できない場合もあり、その場合は、検出できた角の情報のみを利用して、カメラキャリブレーションを行っている。

この自動検出された角の情報やチェッカーボードの辺の長さなどを利用してカメラの内部パラメータや外部パラメータを求めた。LiDAR から得られたチェッカーボードの点群を選択し、内部パラメータと外部パラメータを用いてカメラ座標系へ変換する。そして画像平面上に投影する。そして、画像から求めたチェッカーボードの位置とのユークリッド距離を計算する（再投影誤差）。再投影誤差の平均は 0.73 pixel であった。本研究では、カメラに対して、チェッカーボードが正面になるように撮

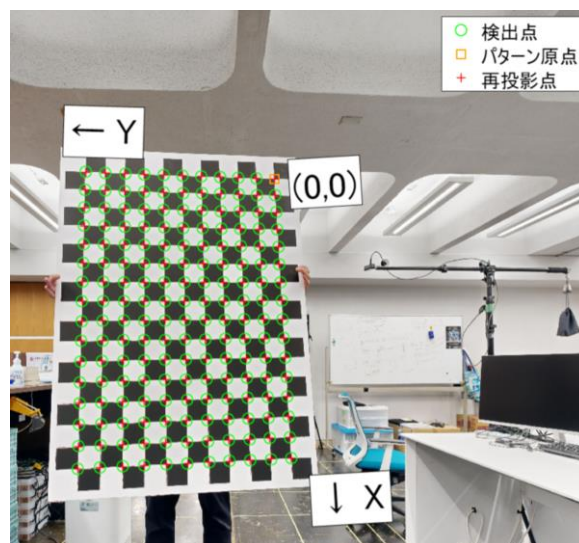


図5 カメラキャリブレーションをしている様子。チェッカーボード中のパターンが自動で検出されている。

影し、それに加え、斜め方向になるようにチェッカーボードを移動させながら撮影を行った。このように多様な種類のチェッカーボードの画像を用意したことが、キャリブレーションをうまく行うことのできた理由の一つであると考えられる。また、本研究では、チェッカーボードの角の位置を自動的に検出しているが、同様の手法を利用する場合は、明るさが安定している場所でチェッカーボードを撮影することや、カメラの画角にチェッカーボードを収めることなどが重要である。明るさが場所によって異なると、チェッカーボードの色味が画像中で変化し、角の検出を精度良く行うことができないという可能性が生じる。しかし、カメラと LiDAR のクロスキャリブレーションの精度が低い場合、対象と LiDAR の距離が遠くなるほど、マッピングの誤差が大きくなる。そのため、さらに高精度なキャリブレーションを行うことも望まれる。利用するチェッカーボードの数を多くし、より多くのサンプルでキャリブレーションを行うことや、多様な状況のサンプルを取得することなどが高精度化のための方策として挙げられる。

#### b) LiDAR とカメラのクロスキャリブレーション

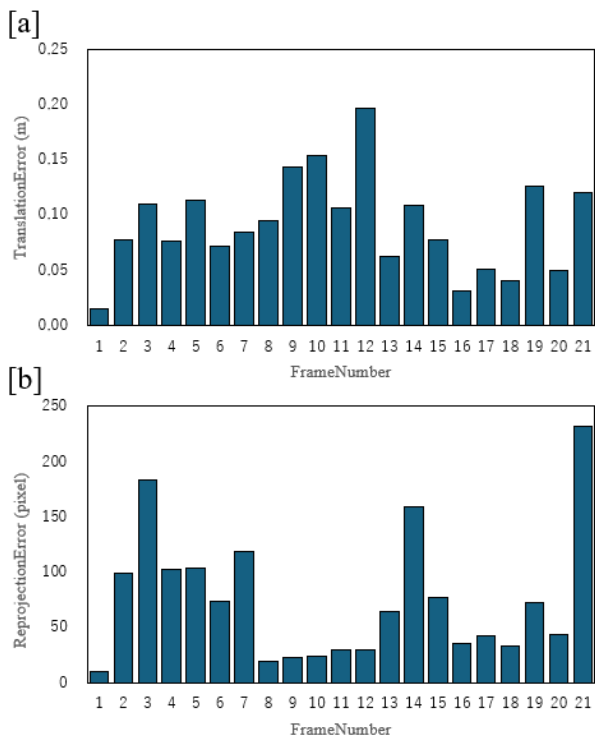


図-6 LiDAR とカメラのキャリブレーションの結果。 [a]は、求めた外部パラメータや内部パラメータを利用して、チェッカーボードの座標（世界座標）をカメラ座標に直したときのカメラ座標との差（メートル）を示す。 [b]は、求めた内部パラメータや外部パラメータをもとにチェッカーボードの点群を画像座標上にマッピングしたときの座標を画像におけるチェッカーボードの座標と比較し、その誤差（ピクセル数）を求めたときの結果。

以下の図-6に、フレームごとのLiDARとカメラのクロスキャリブレーションを行った時の誤差を示す。 [a]は、求めた外部パラメータや内部パラメータを利用して、チェッカーボード点群の中心の座標（世界座標）をカメラ座標に直した座標と、カメラから求めた、チェッカーボードの中心のカメラ座標の差（メートル）を示す。 また、 [b]は、求めた内部パラメータや外部パラメータをもとに、チェッカーボードの点群の中心を画像座標上にマッピングし、その座標と、画像におけるチェッカーボードの中心の画像座標の差（ピクセル）を求めている。これらの値は小さいほど、高い性能を示している。入力には、23組のチェッカーボードの画像とその点群を利用した。しかし、そのうち2組は、画像からチェッカーボードの角の座標を精度良く求めることができなかった。そのため、図-6のグラフでは21組の結果のみを掲載している。画像中にチェッカーボード全体の様子がうまく映せておらず、チェッカーボードの角の部分の情報が欠落していたといったことが原因であると考えられる。

#### c) キャリブレーションの結果を利用した、チェッカーボードの点群のカメラへのマッピングについて

図-7は、点群からチェッカーボードを手動にて切り出し、その点群を画像にマッピングしたときの結果を示している。点群中のチェッカーボードの各点を画像にマッピングしたときの様子を赤色で示している。画像中のチェッカーボードの領域に赤色の点がプロットされていることがわかる。これは、チェッカーボードの点群のxyz座標をもとに、画像中の対応する領域にうまく点をマッピングできていることを示している。このように画像中のチェッカーボードと、点群のチェッカーボードの座標を画像座標に変換した場合の結果が一致しており、LiDARとカメラのクロスキャリブレーションがうまく実行できていることを示唆している。また、キャリブレーションが精度良く実行できていない場合は、図-7の赤色の点が、画像中の別の場所にプロットされ赤色のプロットと画像中のチェッカーボードの位置にずれが発生する。このように、カメラの内部パラメータや外部パラメータを求めることで、カメラの情報や、カメラの光学中心

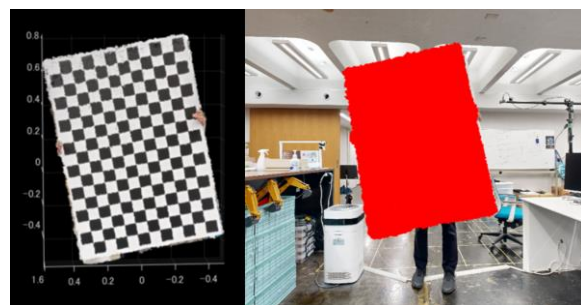


図-7 点群からチェッカーボードを切り出し、その点群を画像にマッピングしたときの結果。

を原点としたカメラ座標と LiDAR 点群の有する世界座標の紐づけを行うことができた。これは、画像のピクセルの位置と点群の各点の対応関係がわかるということの意味しており、この手法を利用することで、カメラ画像中で検出した人の位置を、3次元点群を利用して求めることができることがわかった。

本手法では、チェッカーボードを利用しており、カメラから取得された色画像以外の画像とセンサーフュージョンを行うことも可能である。適用可能な画像の種類が多岐にわたることも本手法の利点である。例えば、サーモカメラの情報と、点群の情報を紐づけることも可能である。しかし、サーモカメラの場合は、対象の色情報ではなく、温度の情報が反映されているため、画像から自動的にチェッカーボードを検出することがより困難になると考えられる。その場合は、手動にてサーモカメラからチェッカーボードの角の座標を特定することや、チェッカーボードの温度が周辺の物体と異なる状況にてデータ取得をするといった対策が考えられる。

本研究での手法の他に、カメラ画像と点群を結びつけるために、オプティカルフローを利用した方法も知られている<sup>20)</sup>。カルマンフィルタも利用しながら頑健なセンサーフュージョンの手法が構築できることが報告されている。センサーフュージョンにおいては、チェッカーボードの座標に加え、人工構造物の角や線といった、チェッカーボード以外の周辺の特徴的な点も加えて、キャリ

ブレーションを行うと、より高精度にセンサーフュージョンを行うことができると考えられる。また、実験フィールドにて繰り返し LiDAR 計測を行う中で LiDAR とカメラのセンサーフュージョンのためのパラメータが微妙に変換する可能性もある。そのような場合に対応するために、パラメータ調整を行う方法も提案されており<sup>21)</sup>、実運用の際はそのような手法も組み込まれることが望ましい。

## (2) 点群からノイズ除去を行った時の結果

図-8は、ノイズ除去を行うために利用した入力画像と、ノイズ除去を行う対象の点群ファイルの例を示す。

[a]は入力画像の1つで、人は含まれていないことがわかる。SOLOv2 を利用し、セグメンテーションを行った場合、人は検出されなかった。[b]も同様に入力画像の1つを示している。この画像では、右側の人が検出され、セグメンテーションされている。この位置で画像を取得した時の点群データを[c]に示す。画像中央部に人が映っている。これは画像[b]の中の人と対応する。[d]は、[b]のセグメンテーション結果を点群にマッピングした時の結果である。人が赤色で塗られており、点群上でもうまくセグメンテーションされていることがわかる。[a]では、人は存在しないため、そのような人のセグメンテーションが実行されなかったことは、正しい推論結果である。また、[b]内では、人が存在し、その領域のみ正しくセ

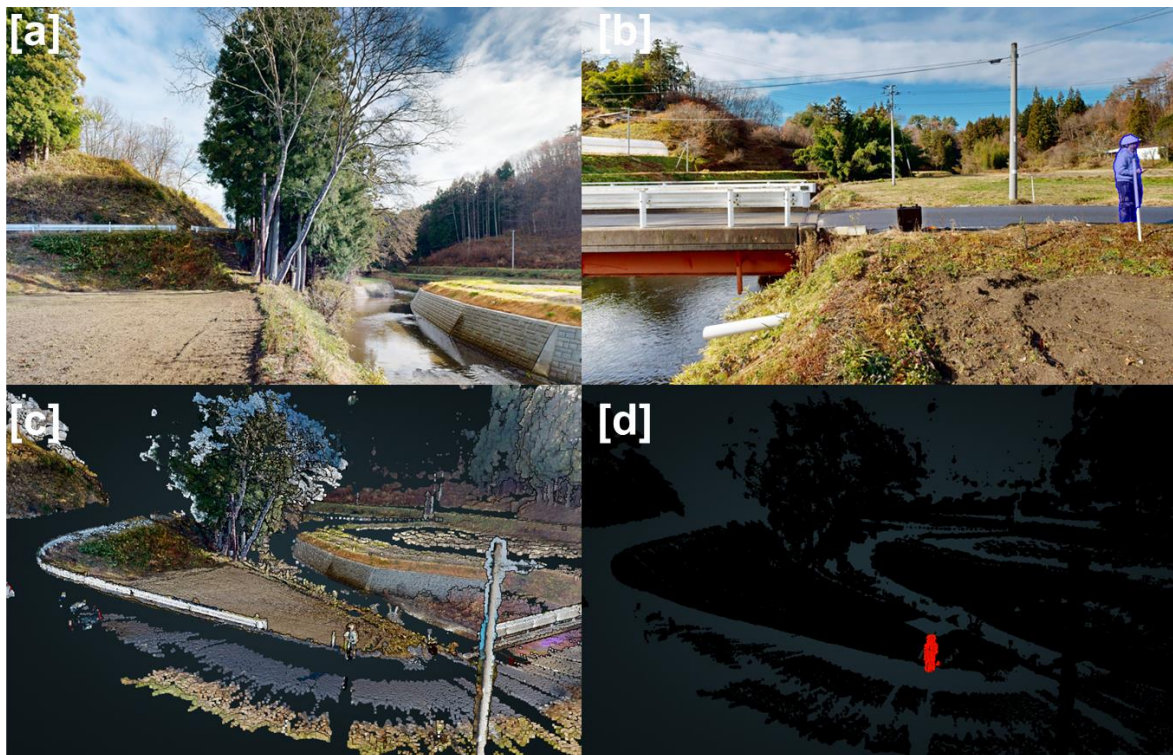


図-8 ノイズ除去を行う時に利用した画像と点群データの例とノイズ除去の結果。[a]には人が存在せず、SOLOv2 によるセグメンテーションの結果人は検出されていない。[b]にはセグメンテーションにより右側の人が検出される。この人に対応する点群[c]に示され、[d]ではこの人のセグメンテーション結果が点群上に赤色でマッピングされている。

グメンテーションされている。そのため、[d]の点群データでも、人以外をセグメンテーションする偽陽性は見られなかった。この理由として、COCO データセットで訓練された、SOLOv2 ネットワークが精度良く人をセグメンテーションできていたことが挙げられる。本研究では、人のセグメンテーションのために、ファインチューニングは行っておらず、学習済みモデルをそのまま利用している。このように、学習済みモデルを利用できる場合は、モデルの構築が効率化される。Segment Anything<sup>2)</sup>などの画像上の物体を精度良くセグメンテーションするその他の技術を利用することで、より複雑な点群上での物体のセグメンテーションも実現する可能性がある。また、本手法を利用することで、人以外のカテゴリの物体もセグメンテーションが可能となる。また、画像の撮影

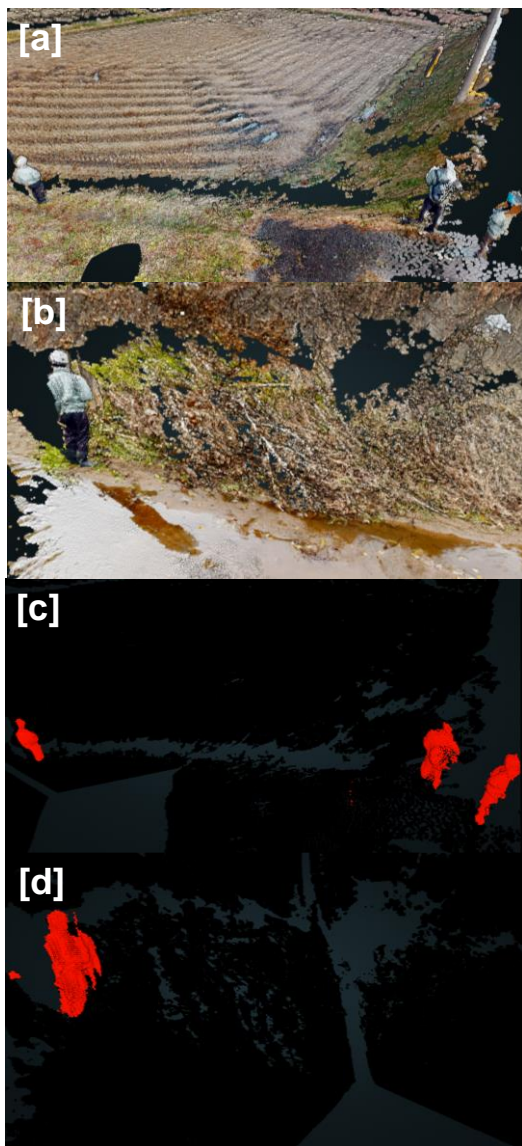


図-9 ノイズ除去を行う際に利用した点群データとそのデータからノイズ除去を行った時の結果の例。[a]および[b]は対象の点群ファイルを示す。また、[c]と[d]はそれぞれ、[a]および[b]から人のノイズ除去を行った時の様子を示す。

角度によって人物の後ろに生じる影に隠れた箇所(point cloud)は、ノイズとして除去されることなく元の点群データに残る。2章(3)で述べた通り、画像から得られたノイズの情報を投影した3次元点群のノイズの点に対してクラスタリングを行い、LiDARから最も近いクラスタのみを保持するように設計しているためである。

図-9は点群データと、その点群から人の分類を行った時の結果の例である。本手法により分類された人の点群は赤く示されている。[a]および[b]は対象の点群ファイルを示す。また、[c]と[d]はそれぞれ、[a]および[b]から人のノイズ除去を行った時の様子を示す。色情報が付与されているため、見た目では人の点群を認識することができるが、どの場所にでも人が立つことができ、場所に制約がないことや、地表面の形状も平面的でないことから、人の位置を点群のみから認識することは容易ではない。しかし、図-8と同様に、以下のように他の物体が多く存在する中でも、人をうまく分類できていたことができた。本手法を利用することで、様々な場所に人が存在する中でも、画像から人を検出することができれば点群でもノイズとして除去することができる。例えば橋梁などの構造物の上を歩いている人物の点群を処理する際、セグメンテーションの範囲が広がるとノイズの過検出が生じることがある。これは使用しているSOLOv2のセグメンテーションアルゴリズムの性能に依存する部分が多いためである。そのため、セグメンテーションが適切に機能しない場合には、構造物の点群が誤って除去されたり、人物の足の点群が部分的に残ることがある。このように画像のセグメンテーションの性能がノイズ除去の精度に直接的に関わってくるため、その課題に合わせたセグメンテーションアルゴリズムを用意することや、ファインチューニングなどの手法により、その状況にあったモデルを学習させる必要がある。

### (3) 点群からノイズ除去を行った時の結果

以下の表-1にテストデータ中の各人の点群のノイズ除

表-1 テストデータ中に含まれる、人のノイズ除去の精度検証の結果。

サンプルID	Recall	Precision	F1
1	0.658	1.000	0.794
2	0.876	1.000	0.934
3	0.966	1.000	0.983
4	1.000	0.899	0.947
5	0.985	0.915	0.949
6	0.957	0.553	0.701
7	0.990	0.960	0.975
8	1.000	0.972	0.986
9	0.934	0.607	0.736
平均	0.930	0.878	0.889



去の精度を示す。各サンプルに対する、RecallやPrecision, F1 の値を示している。すべてのサンプルに対する、Recall および Precision の平均は、それぞれ、0.923 および、0.878 であった。また、それらの調和平均である F1 は 0.889 であった。表-1にて示されたように、本手法の精度を評価した結果、おおむね高い精度を得ることができた。図-10 に表-1 の結果を可視化した時の様子を示す。[a]~[d]の図はいずれも人の輪郭を計測した3次元点群である。側面から計測されている場合があり、さらに LiDAR との距離が遠い状況においては、輪郭が曖昧になっている場合があった。[a]はサンプル ID3 であり、Recall, Precision とともに高かった。赤が本手法にて分類した時の結果、青が、手動にてラベリングした人のノイズの点である。赤と青の点の分布がほとんど一致しており、高い精度となっている。また、図-10 [b]はサンプル ID8 であり、同様に分類精度がよい結果の例である。一方、[c]はサンプル ID1 の結果であり、分類した赤の結果は正解データの一部のみしか網羅していないことがわかる。この理由として、人が計測中に移動したことが考えられる。ここでは、分類した点は全て人であったため、Precision は1であるが、分類されていない点が多数存在したため、Recall は低い値となっている。また、[d]はサンプル ID6 の結果を示している。より広い領域をノイズとして分類したため、赤の範囲が大きい。そのため、Precision は低い値になっている一方で、人の領域はほとんど網羅しているため、Recall は高い値となっている。人の輪郭の情

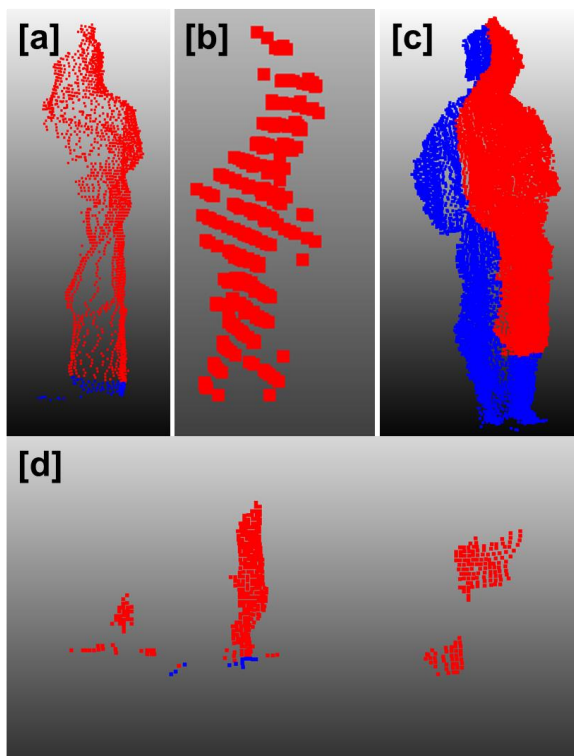


図-10 ノイズ除去の結果の例。赤は本手法により分類した時の結果で、青は手動にてラベリングした結果である。

報を膨張して点群にマッピングしているため、点群に過剰にノイズの点が与えられる場合がある。この場合、膨張のパラメータを小さくすればより高い精度にて分類が可能となる。状況に応じてパラメータ調整を行うことが可能である。また本研究により評価したサンプルには限りがあり、今後の展望としてより多様な状況下にて、より多くのサンプルにて本手法の精度評価をすることが望ましい。例えば、立った状態の人だけでなく、座った人の点群や、走っている状態の人のノイズをうまく分類できるかどうかといったことも検討されることが望ましい。

また、より精度良く画像のセグメンテーション情報を点群にマッピングするために、LiDARとカメラのキャリブレーションを高精度に行うことが望ましい。より多くのサンプルを利用して、多様な状況にてキャリブレーションを行うことで、高精度化が可能になると考えられる。

LiDAR点群から人を検出するために、点群から形状の特徴量を計算し、サポートベクトルマシン(SVM: Support Vector Machine)にて人を検出した研究も存在する<sup>23)</sup>。しかし、そこでは地表面の抽出をし、その後に人を含むクラスターの解析をしている。本研究のデータのような、地表面の形状が複雑であったり、対象と計測位置が離れていたりするようなデータであれば、そのような方法では人の分類は難しいと考えられる。

#### (4) 点群からノイズ除去を行った時の結果

図-11 は人が計測地点から離れている場合の結果につ

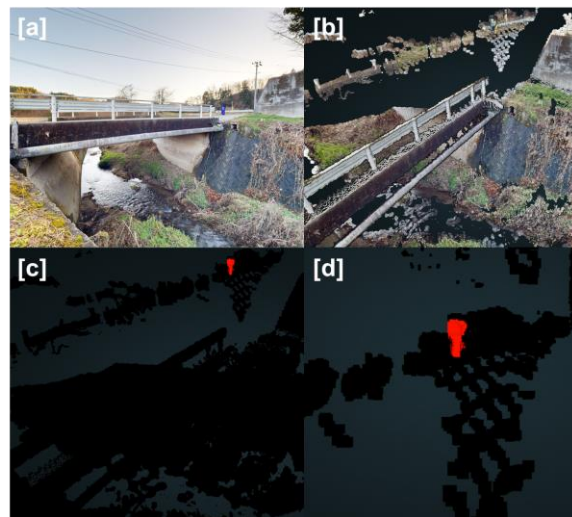


図-11 計測地点から人が離れた状況におけるノイズ除去の結果。この結果においては、画像の膨張処理を施さず、点群にセグメンテーションの結果を直接マッピングしている。[a]は実際の橋の写真であり、橋全体とその周辺環境が含まれている。[b]では橋及びその周辺の点群データが示されており、[a]からの画像がマッピングされ色付けされている。[c]と[d]は、この手法を適用した際の点群データの処理結果を示している。

いて示す。なお、本結果では、画像の膨張処理を行わずに、点群にセグメンテーション結果をマッピングしている。[a]は実際の橋の写真で、周囲の環境も含めて橋全体が見えている。[b]は橋と周囲の環境の点群データを示す。[a]の画像がマッピングされており、色情報が付加されていることがわかる。[c]と[d]は点群データ中で本手法による処理を行った時の結果を示しており、赤い点はノイズとして分類された点である。このように、LiDARから距離の離れた対象であっても、本手法を利用して分類を行うことができた。LiDARの位置から離れた対象は点群の密度が下がるため、物体の認識が難しくなる。しかし、本手法では画像をもとに物体を認識しているため、そのような状況でも点群の分類が可能となる。また、2次元画像にて分類を行っており、対象となるノイズの手前に物体が重なって存在する場合は、その物体がノイズとして分類されるという課題が存在する。LiDARから最も近い物体をノイズとして分類するため、人と、別の物体が重なっている場合は、適切にノイズ除去ができない可能性がある。2次元画像にてセグメンテーションを行う際に、そのような重なりがないかを確認するプロセスを追加することでそのような課題を克服できる可能性がある。また、画像を取得する際、悪天候や明るさが不十分である場合は、対象の様子を画像から捉えることができず、セグメンテーションが困難となる。このような場合に備え、より多様な状況下でうまくセグメンテーションできるモデルを用意することや、その状況に応じた前処理方法などを検討する必要がある。

細井ら<sup>24)</sup>は475-840 nmの波長を入力に取ることができるマルチスペクトルカメラと地上型レーザーキャナから取得された3次元点群のセンサーフュージョンを行った。この研究では、樹木の3次元形状をLiDARにて計測し、マルチスペクトルカメラの情報を点群にマッピングしている。マルチスペクトルカメラの情報は植物において重要な成分であるクロロフィル量と相関があり、その情報を点群にマッピングすることで3次元的なクロロフィルの分布を可視化している。このように、マルチスペクトルカメラやサーマルカメラなどの、色情報以外の情報を取得できるカメラの情報と点群のセンサーフュージョンを行うことで、対象物体の内部品質などの検査にもつながる可能性がある。

また本研究では、地上型レーザーキャナが利用されたが、MMSといった別のレーザーキャナを対象に行うことも可能である。例えば、MMSの計測中に取得した画像から損傷部を検出し、MMSの位置や画像の向きなどを考慮し、損傷の位置を3次元的に把握できる可能性がある。今後は、地上型レーザーキャナ以外の点群の取得方法にも着目し、本手法の適用可能な範囲をより詳細に検討することが望ましい。

## 4. まとめ

本研究では、LiDARとカメラのセンサーフュージョンを用いた3次元点群のノイズ分類が行われた。はじめに、チェッカーボードを使用してカメラとLiDARの相対位置の調整を実施し、その後、画像から点群へのマッピングを可能にする内部パラメータと外部パラメータを求めた。これにより、点群の各点と画像上のピクセルの位置関係の紐付けができ、そのセグメンテーションの手法が可能となった。次に、画像にてセグメンテーションした情報を点群にマッピングを活用して点群のノイズ分類を実施した。具体的には、2次元画像上で人の輪郭を抽出し、それを3次元空間にマッピングすることで、点群内の人に関連するノイズ点を分類した。この手法により、他の物体が多数存在する状況下でも、人を効果的に分類できることが確認された。

Recall, Precision, 及びF1スコアを用いた評価結果は、全サンプルにおける平均値がそれぞれ0.923, 0.878, 0.889であり、高精度なノイズ除去が可能であることが示された。

しかし、本手法では対応が難しい場合や、今後より検討が必要な場合も存在する。例えば、高速で移動する自動車などの物体は、LiDARによる計測中に位置が変動してしまうため、正確なセグメンテーションが困難である。この理由は、画像における対象の位置と点群での位置にずれが生じるためである。そのため、今回のセンサーフュージョンの方法では対応が難しく、別途点群処理による対策を検討する必要がある。また、LiDARでは半透明の物体を計測するのが難しく、不完全な形で点群化されることがある。一方、画像センサーを用いることにより、これらの物体のセグメンテーションが可能である。したがって、半透明物体に対するセグメンテーションの可能性を今後さらに検討する必要がある。

本研究の成果は、3次元点群のクリーニングや前処理に有効であることが期待される。今後の展望として、より多くの人が存在するデータなどで、本手法の精度評価を行うことなどが期待される。

**謝辞：**本稿は、JSPS 科研費 JP21H01417 の助成を受けた研究で行われた結果を含みます。また、内閣府総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム (SIP) 第3期「スマートインフラマネジメントシステムの構築」JPJ012187 (研究推進法人:土木研究所) で実施されている研究も含みます。これらに謝意を表します。

## 参考文献

- 1) 山下淳子, 木村沙智, 川村日成: 3次元点群データを活

- 用したインフラ構造物の維持管理. 精密工学会誌, Vol. 85, No. 3, pp. 228-231, 2019.
- 2) 大伴真吾, 鈴木清, 土橋浩, 永田佳文, 菅野晶夫, 安中智, 乾義文: 道路・構造物維持管理支援システムにおけるポイントクラウドの高度利活用について. 写真測量とリモートセンシング, Vol. 55, No. 1, pp. 27-31, 2016.
  - 3) Chen, Y., Lin, J., and Liao, X.: Early detection of tree encroachment in high voltage powerline corridor using growth model and UAV-borne LiDAR. *Int. J. Appl. Earth Obs. Geoinf.*, Vol. 108, pp. 102740, 2022.
  - 4) Inomata, T., Fernandez-Díaz, J. C., Triadan, D., García Mollinedo, M., Pinzón, F., García Hernández, M., ... and Moreno Díaz, M.: Origins and spread of formal ceremonial complexes in the Olmec and Maya regions revealed by airborne lidar. *Nat. Hum. Behav.*, Vol. 5, pp. 1487-1501, 2021.
  - 5) Itakura, K., and Hosoi, F.: Automatic individual tree detection and canopy segmentation from three-dimensional point cloud images obtained from ground-based lidar. *J. Agri. Meteorol.*, Vol. 74, pp. 109-113, 2018.
  - 6) Itakura, K., Miyatani, S., and Hosoi, F.: Estimating tree structural parameters via automatic tree segmentation from LiDAR point cloud data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, Vol. 15, pp. 555-564, 2021.
  - 7) Itakura, K., and Hosoi, F.: Three-dimensional tree monitoring in urban cities using automatic tree detection method with mobile LiDAR data. *Arti. Intel. Data Sci.*, Vol. 2, pp. 1-10, 2021.
  - 8) 峰岸樹, 江守央, 佐田達典: 点群データに含まれるノイズの統計的・幾何的手法を用いた自動的除去に関する研究. 土木学会論文集 F3 (土木情報学), Vol. 78, No. 2, pp. 49-55, 2022.
  - 9) Rakotosaona, M. J., La Barbera, V., Guerrero, P., Mitra, N. J., and Ovsjanikov, M.: Pointcleannet: Learning to denoise and remove outliers from dense point clouds. *Computer graphics forum*, Vol. 39, No. 1, pp. 185-203, 2020.
  - 10) Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. *ECCV*, pp. 801-818, 2018.
  - 11) 劉佳明, 党紀, 全邦釘. (2022). DeepLabv3+ を用いた橋梁腐食損傷とその精度の向上. *AI・データサイエンス論文集*, 3(J2), 802-810.
  - 12) He, K., Gkioxari, G., Dollár, P., and Girshick, R.: Mask r-cnn. *ICCV*, 2017.
  - 13) Wang, X., Kong, T., Shen, C., Jiang, Y., and Li, L.: Solo: Segmenting objects by locations. *ECCV*, 2020.
  - 14) Wang, X., Zhang, R., Kong, T., Li, L., and Shen, C. Solov2: Dynamic and fast instance segmentation. *Adv Neural Inf Process Syst.*, Vol. 33, pp. 17721-17732, 2020.
  - 15) Yamane, T., Chun, P. J., Dang, J., and Honda, R. (2023). Recording of bridge damage areas by 3D integration of multiple images and reduction of the variability in detected results. *Comput. Aided Civ. Infra. Eng.*, 38(17), 2391-2407.
  - 16) Chu, H., and Chun, P. J.: Fine-grained crack segmentation for high - resolution images via a multiscale cascaded network. *Comput. Aided Civ. Infra. Eng.*, Vol. 39, No.4, pp. 575-594, 2024.
  - 17) Jiao, J., Chen, F., Wei, H., Wu, J., and Liu, M.: Lce-calib: automatic lidar-frame/event camera extrinsic calibration with a globally optimal solution. *IEEE ASME Trans. Mechatron.*, 2023.
  - 18) He, K., Xiangyu Z., Shaoqing R., and Jian S.: Deep residual learning for image recognition. *CVPR*, pp. 770-778. 2016.
  - 19) Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P.: Microsoft coco: Common objects in context. *ECCV*, pp. 740-755, 2014.
  - 20) 齊藤真衣, 沈舜聡, 伊東敏夫. LiDAR とカメラを用いたセンサフュージョンによる遠距離スパース点群の補間手法. *自動車技術会論文集*, Vol. 53, No. 3, 598-604, 2022.
  - 21) Nedeveschi, S.: Online cross-calibration of camera and lidar, *ICCP*, pp. 295-301, 2017.
  - 22) Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... and Girshick, R. Segment anything. *ICCV*, pp. 4015-4026, 2023.
  - 23) 横田隆之, 黒田洋司: LIDAR を用いた形状的特徴による人認識, *ロボティクス・メカトロニクス講演会講演概要集*, 3PI-K04(1)-3PI-K04(4), 2014.
  - 24) Hosoi, F., Umeyama, S., and Kuo, K. Estimating 3D chlorophyll content distribution of trees using an image fusion method between 2D camera and 3D portable scanning lidar. *Remote Sens.*, Vol. 11, No. 18, 2134, 2019.

# Noise Removal from Point Clouds through Sensor Fusion of LiDAR and Camera

Kenta ITAKURA, Takuya HAYASHI, Yuto KAMIWAKI and Pang-jo CHUN

In this study, the method of noise removal in 3D point clouds using sensor fusion of LiDAR and camera was introduced. First, a point cloud measurement was performed using Matterport Pro3 in Fukushima Prefecture, Japan. The bridges and other man-made objects were scanned while people were also scanned. The relative positions of the camera and LiDAR were adjusted using a checkerboard for cross-calibration. Internal and external parameters were then obtained to map the 2D image to the 3D point cloud. This allowed us to correlate each point in the point cloud with a pixel in the image and develop a segmentation method for noise removal. Next, we extracted the region of a person in a 2D image and mapped it to a 3D space to classify noise points related to the person in the point cloud. It was confirmed that this method can effectively classify people even in the presence of many other objects in the point clouds. The evaluation results using recall, precision, and F1 scores had mean values of 0.923, 0.878, and 0.889 for all samples, respectively, indicating that highly accurate noise reduction is possible. The results of this study are expected to be effective for cleaning and pre-processing 3D point clouds. Future perspectives include evaluating the accuracy of this method on data with a larger number of people.