三次元深度センサーを用いた移動制約者検出手法の提案と評価

右京 莉規[1] 扇田 幹己[1] 山田 遊馬[1] 廣森 聡仁[1] 山口 弘純[1] 東野 輝夫[1]

概要：
健常者と同速度での移動や反応が困難な移動制約者のストレスフリーな移動環境に向け，公共・商業施設等ではエレベータの優先利用や多目的トイレの設置といったバリアフリー化に向けたサービスや設備設置が推進されている．しかし，移動制約者の施設訪問数や行動は正確に把握されていないため，例えば設備設計が必要十分かの定量的検証などが容易に行えないといった課題がある．本研究ではプライバシーに配慮したセンシングが可能な三次元深度センサーを用い，移動制約者を検出可能なエッジ・クラウド連携型の人流検出手法を提案する．エッジデバイスでは，背景差分法とクラスタリングを適用して移動物体に対応するセグメントを三次元点群から高速で検出し，それらをつなぎ合わせて移動軌跡にするとともに，移動制約者を簡易判定するために事前に定義した特徴量を用いて，対応するセグメント検出を行う．クラウドサーバーではセグメントに対して深層学習ベースの手法 PointNet を適用し，属性判定を行う．この際，エッジからクラウドに送信するデータ量とクラウドでの深層学習アルゴリズムの実行負荷を抑制するため，対象者のセグメント時系列から PointNet 判定に最も相応しい撮影角における 1 セグメントのみを抽出し，クラウドに送信するアルゴリズムを設計している．大型商業施設のエントランスにおいて 7 時間にわたり収集した人流データを用いた検証の結果，エッジデバイスの処理が 271ms, クラウド側での PointNet 判定を含めた処理時間も 275ms で実行できた．またベビーカー利用者判定における適合率は 0.818，再現率は 0.667，f 値は 0.735 であった．さらに，エッジでクラウドに送信するデータを選択することにより，データ通信量を約三分の一に抑制することができた．これにより，エッジ・クラウドで適切に負荷分散と通信量削減を行いながら効率良く人流検出が可能なシステム実現の目途を得た．

1. はじめに

国土交通省では，「交通行動上，人の介助や機器を必要としたり，さまざまな移動の場面で困難を伴ったり，安全な移動に困難であったり，身体的苦痛を伴う等の制約を受ける人々」を移動制約者として定義しており [1]，ベビーカーや車椅子利用者，歩行補助杖の利用者などが相当する．

日本は大規模災害が他国と比較して多く発生するが [2]，大規模災害時では迅速な現場避難や，逆に一時避難施設での滞在が必要とされる場合も多い．そこで駅やショッピングモールなどの施設では移動制約者の人数や属性を常時把握しておくことで，施設員による迅速な避難誘導や避難補助，常備すべき器具数の見積もりなども可能となる．また，駅などの公共施設ではバリアフリー化が推進されており [3]，施設を訪問する移動制約者数と属性が把握できれば，移動制約者数やその行動を考慮した施設・設備設計となっているかの定量評価も可能となる．

通行者およびその属性を把握する方法としては，映像を利用した技術やシステムが多い．特に近年のオブジェクト認識技術の発展により，RGB 画像から YOLOv3 [4] や R-CNN [5] といった技術を活用し，オブジェクトを検出する手法が多く提案されている．しかし，取得画像には通行者の顔や服装といった個人情報に加え，人流検出に不必要なオブジェクトも同時に映り込んでしまう可能性があり，設置角や方向，取得したデータの保護には十分な留意が必要となる．したがって，設置個所や状況によってはプライバシー侵害のリスクが高く，許容されない場合も多い．

近年，対象物や空間の立体的形状を把握できる三次元深度センサーが注目を集めており，人物の骨格情報取得技術 [6] や屋内外の 3 次元地図生成技術 [7] などに用いられている．得られる三次元点群データは物体の形状以外の情報を持たないため，RGB 画像に比べてプライバシー侵害のリスクは低いものの，一方で属性判定のためには 3 次元点群で表現される部分的な形状情報のみから判定を行う必要が

[1] 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

ある．しかし，最近の深層学習技術の飛躍的な発展により，三次元点群データからの対象物を特定に深層学習を適用する研究が進んでおり，特に，三次元点群に対する最先端の物体認識技術である PointNet [8] では，ModelNet40 [9] データセット内の 40 種類の物体識別を高い精度で実現している．こういった手法を用いることで，三次元深度センサーからの属性判定の実現が期待できる．

しかし，その実現には以下の 2 つが課題となる．まず第一に，現場に設置された三次元深度センサーのデータをすべてクラウドサーバーに送信し，データ処理を行うことは通信量を考慮した場合現実的ではない．一方で，エッジデバイスで利用可能な計算資源は限られており，オブジェクト検出や深層学習アルゴリズムによる属性判定などをすべてエッジデバイスで実行することも現実的ではない．これらを考慮し，ある程度現場でデータ処理を行える，エッジデバイスとクラウドサーバーの連携型のアーキテクチャにおいて，クラウドとエッジに適切な機能配分を行い，高い検出精度を実現しながら，エッジとクラウド間の通信量を抑制することが望まれる．第二に，深層学習ベースのアルゴリズムでは，移動制約者の特徴を捉えた適切なデータセット構成を行う必要がある．移動制約者が適切にアノテーションされた三次元点群データを多数得ることは簡単でないため，撮影方向等に多様性のない，数量の限られたデータセットから十分な精度で移動制約者を検出可能なモデル訓練が必要となる．

本研究では，三次元深度センサーを利用した通常歩行者のトラッキングシステムならびにその中から移動制約者を効率よく検出するエッジ・クラウド連携型アーキテクチャを提案する．提案手法では，三次元深度センサーから得られる三次元深度データに対して，エッジデバイスで背景差分法により空間内の移動物体の三次元点群のみを抽出する．そしてクラスタリングに基づくノイズ除去と個々の移動物体（歩行者および移動制約者）のセグメント化を行い，それらを時系列でつなぎあわせることで個々の移動物体のトラッキングを行う．また，与えられたセグメントに対し，それがベビーカー利用者である可能性を簡易判定する特徴量を事前に定義し，クラウドサーバーに配置した PointNet に与えることで属性の最終判定を行う．この際，エッジからクラウドに送信するデータ量とクラウドでの深層学習アルゴリズムの実行負荷を抑制するため，対象者のセグメント時系列から PointNet 判定に最も相応しい撮影角における 1 セグメントのみを抽出し，クラウドに送信するアルゴリズムを設計している．PointNet ではベビーカー利用者, 歩行者， 子供を抱える人, その他の 4 属性を判定できるよう，事前に収集したデータセットを用いたモデル学習を行っている．

大型商業施設のエントランスにおいて 7 時間にわたり収集した人流データを用いた検証の結果，エッジデバイスの処理が 271ms, クラウド側での PointNet 判定を含めた処理時間も 275ms で実行できた．またベビーカー利用者判定における適合率は 0.818，再現率は 0.667，f 値は 0.735 であった．さらに，エッジでクラウドに送信するデータを選択することにより，データ通信量を約三分の一に抑制することができた．これにより，エッジ・クラウドで適切に負荷分散と通信量削減を行いながら効率良く人流検出が可能なシステム実現の目途を得た．

本稿の構成は以下の通りである．2 章では，関連研究として，三次元点群を活用した対象物のコンテキスト推定手法について紹介し，それらに対する本研究の位置付けを明確にする．3 章では，提案手法の概要と，本研究における移動制約者の定義を述べる．4 章では，提案手法の詳細について説明し，5 章では，ショッピングモールで実施した実験とその評価について述べる．最後に 6 章で本研究のまとめと今後の課題について述べる．

2. 関連研究

近年，三次元深度センサーの普及に伴い，センサーにより得られる三次元点群データを対象に，物体の種類や人の骨格などのコンテキストを推定する研究が多く為されている．文献 [8] では，ディープニューラルネットワークにより，入力として与えられた三次元点群データに対し，その物体が何であるかを判定するクラス分類，ある物体を個々の部位に分割するセグメンテーション分類を実現する PointNet が提案されている．この手法においては，三次元点群データの密度が均一であることを仮定しているが，実世界おいて計測される三次元点群データの密度は必ずしも均一ではないため，学習した三次元点群データと異なる密度の三次元点群データに対する認識精度は低下する．文献 [10] では，より現実的な環境において，物体を認識するために，異なる粒度の三次元点群データを扱えるよう，前述の PointNet を階層的に適応することにより，様々な物体が配置された複雑な状況においても，物体を認識できる手法が提案されている．また，VoteNet [11] では，三次元点群データに対し，ハフ変換を活用した投票により，物体の中心位置を予測し，これに基づき，物体に対するバウンダリボックスを高い精度で導出する手法を提案している．同様に，文献 [12] においても，三次元点群データ中の物体に対するバウンディングボックスを予測することにより，個々の物体に紐づく三次元点群データを抽出する手法が提案されている．一方，物体を認識だけでなく，人の姿勢を把握する手法がいくつか提案されている [6, 13–15] Complex YOLO [14] は，取得し三次元点群データ群を俯瞰して見た場合の二次元画像に変換し，その画像に対し物体検出の手法を適応することで，物体の存在を把握する手法である．文献 [13] においては，点群から手や体の姿勢を推定するために，三次元点群データを二次元深度画像に変

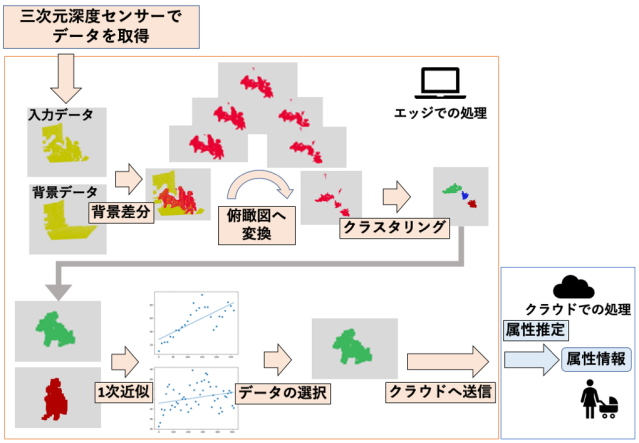


図 1　ベビーカー利用者属性推定フローチャート

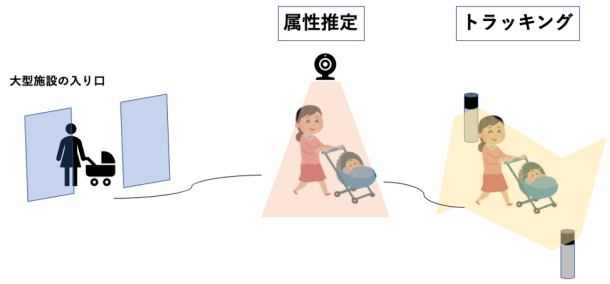


図 2　想定する動作シナリオ

換する手法と異なり，三次元点群データを 3D ボクセルに変換することにより，個々の形状を特徴を維持し，これに基づき推定する手法を提案している．また，評価実験においては．人を前面と上面から撮影した，いずれの場合においても，高精度に全身の姿勢を推定できることを示している．文献 [16] では，入力となる三次元点群データに対し，予め定義された姿勢のうち，尤もらしい姿勢を高速に導出する手法を提案している．また，慣性センサーと単一視点深度カメラからの情報を融合することにより，高速な動きに追随し，リアルアイムに姿勢を推定する手法も提案されている [15].

これらの深層学習による手法の多くは，高い精度で物体や姿勢を認識できるよう，GPU を伴う計算機を必要としているものがほとんどである．一方，本取組では，移動制約者の存在を把握するための一連の処理を，エッジとクラウドの双方に分散するものであり，様々な場所で利用できる手法となっている．

3.　システムアーキテクチャと動作概要

本研究は，三次元深度センサーから記録される三次元深度データを解析することによって，対象領域内を移動する人が移動制約者かどうかを推定することを目的としている．本システムでは，対象としている移動制約者のうちベビーカー利用者かどうかの推定を対象としている．

本研究の概要を図 1 に示す．本システムは実際に取得した三次元深度データをエッジで処理し，ベビーカー利用者の可能性の高いセグメントをクラウドに送信し属性を推定する．属性の推定を行うため，三次元点群データから移動制約者の属性を判定するためのモデルをあらかじめ構築しておく．

まず．実際の三次元深度データを実際の座標系に変換し，背景差分法により移動物体の三次元点群データを抽出する．その後抽出したデータを俯瞰視点のデータに変換しクラスタリングを行い，ノイズ除去と個々の歩行者のセグメント化を行う．そして，それぞれのセグメントに含まれるポイントの高さから 1 次近似を行うことで，ベビーカー利用者の可能性の高いセグメントを抽出する．そして，抽出された歩行者の点群データを三次元点群移動制約者属性推定モデルに入力し，その属性を推定する．

三次元点群移動制約者属性推定モデルでは，推定対象となるセグメントを入力すると，推定される属性を出力する．モデル作成のために，実際に大型商業施設のエントランスに三次元深度センサーを設置し取得した三次元点群データから，属性推定対象となる移動物体の三次元点群に対して手動でアノテーションを行い，学習データを作成した．作成したデータと PointNet [8] と呼ばれる深層学習ベースの教師あり学習を用いて，三次元点群データの属性を推定するモデルをクラウド上に構築する．

本システムの想定動作シナリオを図 2 に示す．三次元深度センサーを施設の入り口やエレベーターなどに設置し，来訪者の三次元深度データを取得する．対象領域に対して，一台の三次元深度センサーを活用することを想定しているため，一つの対象物に対して同時に取得できる三次元深度データは単一方向から捕捉したもののみとなる．そして，取得された三次元深度データを三次元点群に変換し，対象となる人物の属性情報をリアルタイムに推定する．また，推定した属性を既存のトラッキング技術に結合することで，属性付加をした人物軌跡を取得することができ，移動制約者のトラッキングを行うことが可能となる．

4.　人流検知と移動制約者の属性判定

本章では三次元深度センサーの対象領域内に存在する移動体のトラッキングとそれらの属性推定方法について述べる．

まず，センサーから取得したデータを 4.1 章で述べる背景差分を行い，移動物体を抽出する．次に，4.2 章で述べる自動クラスタリングを行い，移動物体をクラスタリングする．取得したデータが人流のないデータの場合，そのデータを棄却する．その後，クラスタリングした移動物体から

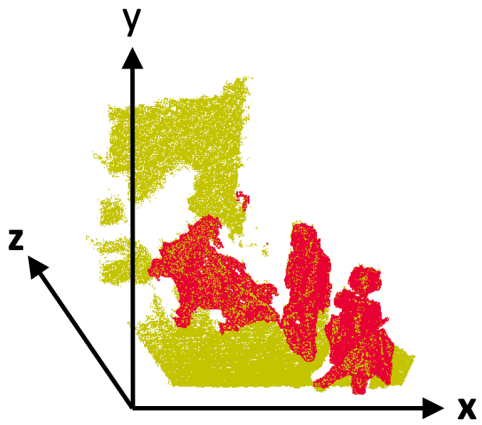


図 3　座標軸の定義

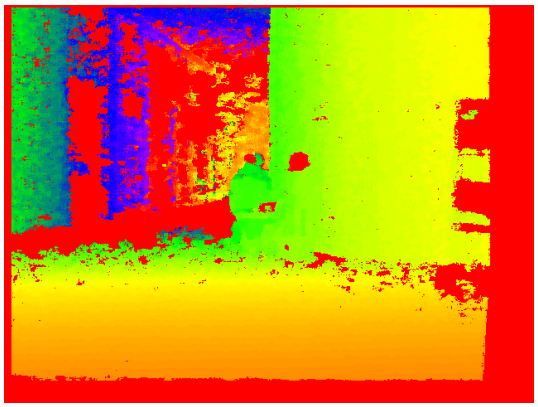
ベビーカー利用者の可能性の高いセグメントのみを，4.3 章で述べる手法で選択する．最後に，4.4 章で述べる手法で選択された移動物体の三次元点群から，対象の属性を推定する．

また，本論文で扱う三次元点群は図 3 の座標軸で表現する．本研究で用いるデータは，三次元深度センサーを x 軸に対し水平に設置して取得したものとなっている．

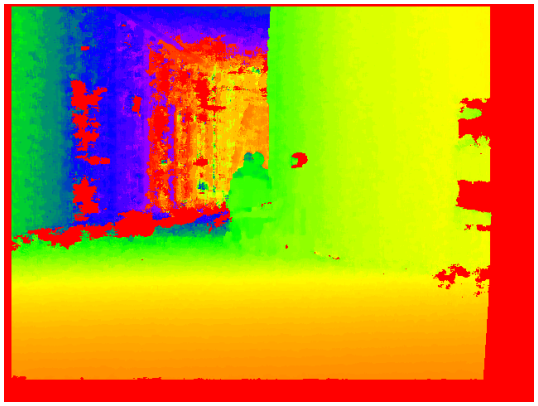
4.1　背景差分による移動物体の抽出

本システムでは，背景差分を行うことで移動物体を抽出している．本システムで行う背景差分は以下の通りである．まず，センサーの対象領域に対して，予め移動物体が存在しない時の，空間の背景となる三次元深度データを取得しておく．深度センサーが取得する三次元深度データには，デバイスに依存した欠損値がフレームごとにランダムに含まれるため，背景データの生成には 30 フレームの三次元深度データを用い，各ピクセルの中央値を取ることでデバイスに依存した欠損値を除去する．図 4(a) は 1 フレームから生成した背景データ，図 4(b) は 30 フレームの中央値から生成した背景データを表しており，赤色で示されている部分が欠損値であるが，30 フレームの中央値を取ることで欠損値の影響が抑えられていることが確認できる．

そして，移動物体が対象領域内に進入したときに，得られる三次元深度データと背景となる三次元深度データとの間に差分が生じるため，その差分が生じた領域のみを抽出することで，移動物体の三次元深度データを取得できる．実際には，三次元深度センサーが取得する三次元深度データにはわずかではあるが誤差を含むため，本研究では，背景となる三次元深度データとの差分が 10cm 以上ある場合のみ移動物体として判定する．また，背景差分によって移動物体として判定された点群データには，デバイスの測定誤差によって生じるノイズが含まれる．そのノイズについては，クラスタリングを行う際に除去を行う．



(a) 1 フレームのみからなる背景データ



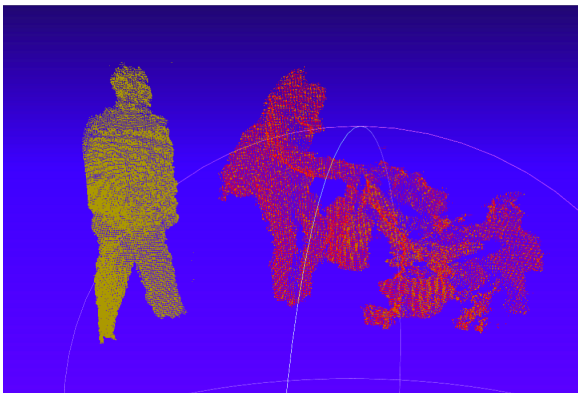
(b) 30 フレームの中央値からなる背景データ

図 4　背景データ

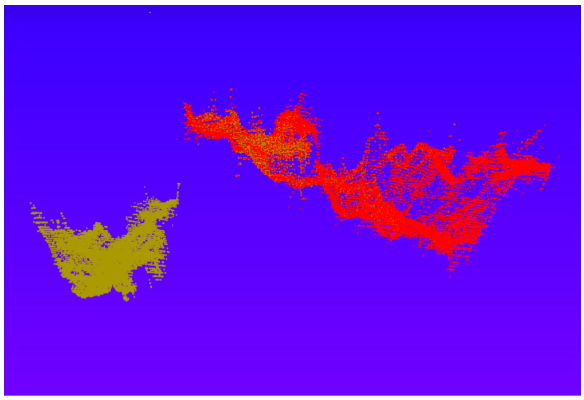
4.2　クラスタリングによる移動体セグメントの抽出

背景差分を行うことによってフレーム内に残った点群データを自動クラスタリングすることにより，ノイズを除去し複数の異なる移動物体を切り分け，それぞれの点群データを抽出する．自動クラスタリングではまず，背景差分法を適用し背景データを取り除くことによって得られた点群データに対し，センサーの仕様で正確なデータを取得できる距離より奥のデータを削除し，ノイズを除去する．本システムで用いたセンサーでは 5m が正確なデータを取得できる最大距離であったため，5m よりも遠くのポイントのデータを削除した．次に処理時間を短縮するため，残っているポイントのうち 1024 ポイントをランダムに抽出しダウンサンプリングする．この時点で残っているポイントの数が 1024 ポイント以下の場合は誰も映ってないフレームとして棄却する．

その後，地面に対して垂直方向の軸を取り除くことで三次元点群を二次元点群に圧縮する．そして，圧縮した二次元平面を 1m 四方の正方形に分割し，それぞれの正方形の中の点が 50 ポイント以上ある正方形とその周囲の 8 つの正方形のみの点を抽出し，その他の正方形に含まれる点を消去する．背景差分後のデータでは，ノイズと移動物体の点群の密度が大きく異なるため，正方形内に点が少ない場合はノイズであるとみなすことができる．また，クラスタ



(a) 横視点



(b) 上視点

図 5 クラスタリングで得られた点群のセグメント

リング時の誤検出をあらかじめ防ぐことができる．

クラスタリング対象となる点群がある場合，得られた二次元点群に対して，DBSCAN アルゴリズムを用いたクラスタリングを適用し，ノイズの除去および，対象となる移動物体の切り分けを行う．三次元点群に対してそのままクラスタリングを適用する方法も考えられるが，システムの対象領域内における移動物体は二次元平面を移動する人，あるいは人によって動かされるもの (ベビーカーや台車など) であり，地面に対して垂直方向の重なりを考慮する必要はないため，二次元平面でのクラスタリングを行う．これにより，クラスタリング処理の高速化も期待できる．また DBSCAN アルゴリズムは，点群の密度に基づきクラスタリングを行うアルゴリズムである．データ内の移動物体は形状や大きさ，数が不定であるが，DBSCAN アルゴリズムはそういったデータに対する堅牢性があり，あらかじめこれらの情報を設定することなく移動物体のデータを抽出することが可能となる．図 5 はクラスタリングを適用し，切り分けが行われた対象物を表している．

4.3 ベビーカーセグメントの簡易抽出

クラスタリングを行った後，セグメントごとにベビーカー利用者であるか否かの簡易判定を行い，クラウドサーバーにおける PointNet 判定の対象とするかを決定する．提案手法では移動体の高さ情報が取得できることを利用し，押している人の身長とベビーカー部分の高さの差，およびベビーカーの高さに現れる特徴に着目してベビーカー利用者の可能性が高いセグメント（候補セグメントとよぶ）を抽出する．ここで，前者の高さの差は，ベビーカー部分の点群の y 軸成分の平均値よりも押している人の点群の y 軸成分の平均値が大きいことで判断する．また，後者のベビーカーの高さの特徴とは，ベビーカーを x 軸の正方向に押している場合，セグメントにおける x の値ごとの y 軸成分の平均値が，x 軸正の方向から負の方向にかけて増加する傾向を示すことである．

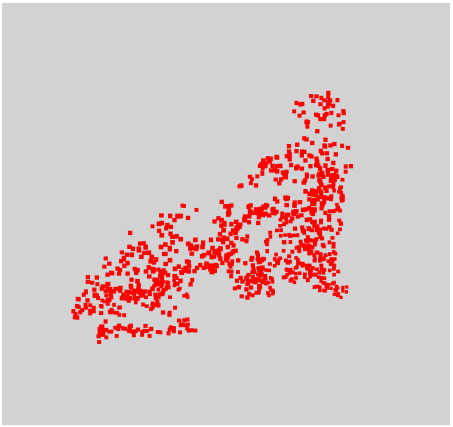
具体的には以下のアルゴリズムで求める．まず，セグメントに含まれる x 軸成分の値が近いポイントごとに y 軸成分の平均値を求める．x 軸成分を大きくした際に，求めた y 軸成分の平均値が上昇傾向もしくは下降傾向のある場合に候補セグメントであると判定した．そのために，セグメントに含まれるポイントを x 軸成分ごとに取得し y 軸成分の平均値を求める．x 軸成分の小さいものから順に求めた平均値を並べる．そしてそれらに 1 次近似を行い，導出された関数の係数をもとに候補セグメントかどうかを判定する．

そのために，まず，x 軸負の方向から順に，セグメントに含まれるポイントの内の 10 ポイントずつで y 軸成分の平均値をとる．その後，求めた平均に対し 1 次近似を行い，近似式の係数を求める．1 次近似では，横軸は x 軸成分の小さいものから順にセグメント内のポイントを並べた際の順番であり，縦軸には y 軸成分の平均値を設定している．横軸の最大値がセグメントに含まれるポイント数となるため，この近似式の係数はサンプル数に影響を受けやすい．この近似式の係数はサンプル数が小さい場合に大きく，サンプル数が大きい場合に小さくなりやすい．この影響を軽減するため，係数に対しセグメントのサンプル数を掛けた値を用い候補セグメントを選択する．本研究では，係数にサンプル数を掛けて求められる数値が一定値よりも大きい場合および一定値よりも小さい場合に，候補セグメントであると判定する．

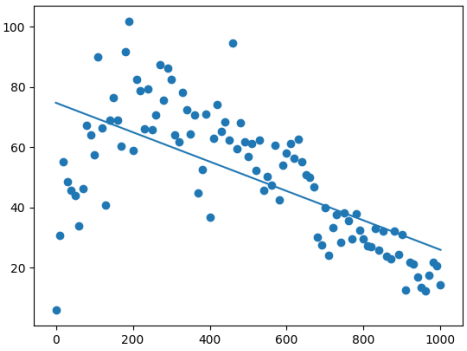
図 6 は，実際の近似の図である．図 6(a) のセグメントから y 軸成分の平均値を取り近似を行うことにより，図 6(b) のような直線が算出される．この近似直線の係数に，このセグメントのポイント数を掛けた数と一定値を比較することにより，候補セグメントかどうかを判定する．そして，ここで選択したセグメントのみをクラウドに送信して PointNet で属性推定する．

4.4 属性推定

本研究では PointNet を用いて移動制約者の属性推定を行う．PointNet を用いた対象の属性推定モデルをあらかじめ構築しておく．そのモデルにエッジで選択したセグメントを入力することにより，推測される属性を出力すること



(a) 候補セグメント



(b) セグメントの長さ平均の 1 次近似を行ったグラフ

図 6 セグメントの長さ平均の 1 次近似

ができる．

PointNet とは三次元点群データを入力として，対象物のクラス分類を行うディープニューラルネットワークのことである．PointNet では，入力点群に対して全結合ネットワークを複数回繰り返すことで，各点に対する特徴量を抽出し，その後，maxpooling 層によって点群全体の特徴を取得する．また，maxpooling 層に対称関数を応用することで，点群の入力の順序に依存しないクラス分類結果を出すことができる点も PointNet の特徴である．PointNet と同様に三次元点群から対象物の属性を推定する手法は数多く存在するが，PointNet では特に，T-Net と呼ばれるネットワークを組み込むことで，点群の回転による影響をなくすことに成功しており，我々のシステムでは対象物の点群データがどの向きから捕捉できるかが事前に把握できないという特性を持つことから，本研究においては PointNet を採用する．

4.4.1 学習データの作成

本研究で作成する PointNet を用いた移動制約者を推定するモデルの学習データの作成について述べる．

本研究では実際に三次元深度センサーを用いて三次元深度データを取得し，“ベビーカー利用者”，“子供を抱える人”，“歩行者”，“その他” の点群データセットを作成した．

まず実験環境にて習得した三次元深度データを基に背景データを作成し，移動制約者や歩行者を捕捉している三次元深度データを入力データとし背景差分を行うことで，三次元深度データから移動物体のみを切り出す．本システムは施設への来訪者の属性推定を目的としているため，捕捉された点群データから背景差分を行うことで移動物体の点群を抽出することができる．そして．背景差分を行うことで得られた点群に対して手動でクラスタリングおよびアノテーションを行う．アノテーションを，データ数が学習に用いるデータの数に到達するまで行い，作成したデータを学習データとして用い PointNet のモデルを作成する．背景差分については 4.1 章で述べた方法を用い，アノテーションについては 4.4.2 章で述べる．また，“その他” については 4.4.3 章で述べる方法を用いて学習データを作成する．

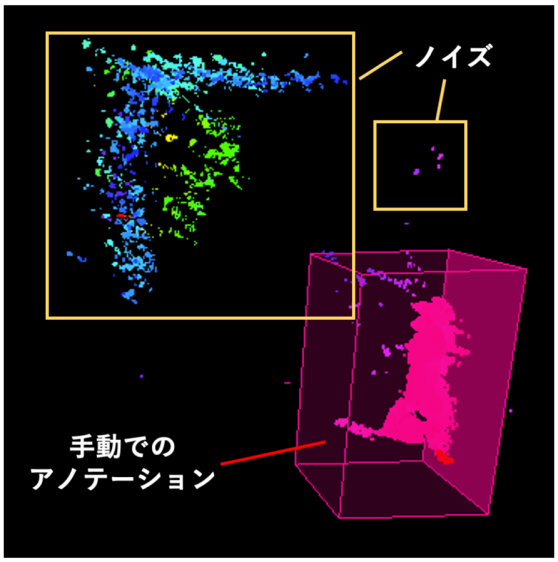


図 7 学習用データの抽出

4.4.2 手動クラスタリング

背景差分を行うことで得られた点群に対し，データのラベル付けが可能な Web サービス Supervisely [17] を用い，手動で移動制約者と歩行者のデータに対してクラスタリングおよびアノテーションを行う．具体的には図 7 のように，移動制約と歩行者などの三次元深度データに箱状のラベル付けを行い，それを基に学習用データを抽出している．また，背景差分によって得られるフレーム内の点群には, デバイスに依存したノイズが多く含まれる．しかし Supervisely を用いたアノテーションは手動で行い，図 7 のように箱状に囲われた部分のみを抽出するため，ノイズに対して堅牢である．

また使用する三次元深度データは，全身を取得できていないデータの学習による誤判定を防ぐため，歩行者は人の全身を取得できているデータのみとする．移動制約者で移動時に道具を用いている場合は，全身および用いている道具の形状も取得できているデータのみを使用する．

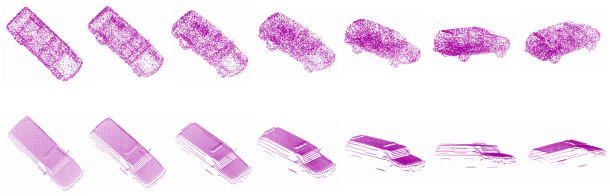


図 8　上段:完全な三次元構造を持つ ModelNet40 の “car” データ，下段:ModelNet40 の “car” データを単一視点から得られる点群に変換したデータ

4.4.3　ModelNet40 データセットの単一視点への変換

本研究で作成した推定モデルには，“ベビーカー利用者”，“子供を抱える人”，“歩行者”，“その他” の四種類を推定できるものを作成した．“ベビーカー利用者”，“子供を抱える人” および “歩行者” については自身で取得したデータを用い，“その他” については ModelNet40 [9] データセットを用いる．

本研究では，単一方向から捕捉される対象物の三次元深度データを用い属性の推定を行うが，ModelNet40 内のデータセットは全て物体の完全な 3D 構造を持つ．そこで，我々は ModelNet40 データセットをそのまま用いるのではなく，ModelNet40 データセットの三次元物体データを三次元深度センサの取得の様子を模した単一視点から観測できる点群のみを切り出し，学習および推定に用いた．具体的には，三次元物体データを三次元深度センサーの画角に収まる位置に配置し，深度画像の各画素について，当該画素の位置に基づいた仰俯角・方位角方向に直線を描画する．そして，その直線と三次元物体が初めて衝突した地点と三次元深度センサーとの距離を当該画素が得る三次元深度データとすることによって ModelNet40 のデータセットを単一視点から取得可能な点群データに変換している．図 8 上段は変換前の完全な 3D 構造を持つ ModelNet40 データセット内の “car” データをあり，図 8 下段は元のデータを単一視点から取得可能な点群データに変換したものである．

5.　評価実験

提案する属性推定モデルの構築と候補セグメントを自動選択するシステムの評価を行うため，実際に大型商業施設のエントランスに三次元深度センサーを 2 台設置し，7 時間にわたり通行データを収集した．その後，三次元アノテーションツールを用いて学習データの作成および移動制約者推定モデルを構築し，評価を試みた．

まず，移動制約者や歩行者の三次元深度データを取得するために行った実験の環境について述べる．本研究では，学習や評価に用いる移動制約者や歩行者の三次元深度データを実際のショッピングセンターで取得した．用いた三次元深度センサーは Occipital, Inc. の Structure Core [18] である．表 1 に Structure Core の性能を示す．

表 1　三次元深度センサーの性能

項目	性能
計測可能距離	0.3 - 5m (最大 10m)
精度	±0.29%
解像度	1280 × 960
フレームレート	54 FPS
視野角	59° × 46° × 70°
消費電力	2.0W (通常時), 3.1W (最大)

表 2　エッジとして利用した計算機の性能

項目	性能
OS	macOS Catalina 10.15.4
PC	MacBook Pro (13-inch, 2019)
CPU	2.8 GHz クアッドコア Intel Core i7

三次元深度センサーは，施設の入り口と店内の間を通過する人のデータを取得できるよう，入り口と店内を行き来するための通路に向くよう設置し，三次元深度データを取得し，移動物体の抽出や属性の判定を行った．また，本システムで，入力データからエッジで処理に用いた PC の性能を表 2 に示す．また，属性推定モデルを用いた属性推定の評価には，GPU として GeForce GTX 1080 を搭載した計算機を用いた．

5.1　評価結果

まず，エッジで推定を行うときの精度について評価を 5.1.1 章から 5.1.3 章で行う．また，作成した属性推定モデルの精度評価を 5.1.4 章で行う．

エッジにおける推定の評価では，実際に取得したデータのうちベビーカー利用者を含む約 2 分間のデータを用いて評価を行った．そのデータは 588 フレームで構成されており，通行した人数は 25 人であった．また，1 フレームに含まれる人数は 0 人から 5 人であった．このシステムの属性推定精度について，処理時間，データ量，分類精度の 3 つの観点から評価する．

5.1.1　処理時間の評価

センサーで通行者のデータを取得してから属性の推定が完了するまでの時間を評価する．まず，エッジでデータを選択する処理時間を図 9 に示す．図 9 の横軸は 1 フレームに対して含まれるセグメントの数を，縦軸にはセグメント数毎の処理時間の平均を示している．セグメント数が 0 の場合処理時間が最も短い 237 ミリ秒である一方，セグメント数が 4 の場合の処理時間が最も長く 303 ミリ秒となっている．このことから，多くのデータに対して 300 ミリ秒程度で処理を実行できていることがわかる．また，1 フレームに含まれるセグメントの数が増えることにより，エッジにおける処理時間が長くなっていることがわかる．これは，セグメントごとにクラウドに送信するかの選択を行うため，セグメント数が多いほど処理にかかる時間が長くな

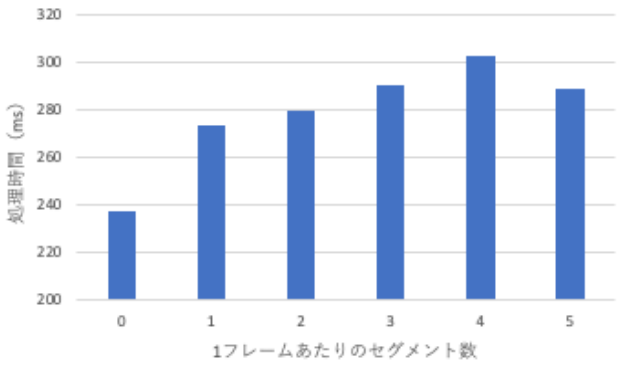


図 9 エッジにおけるセグメント抽出と候補セグメント選択の処理時間

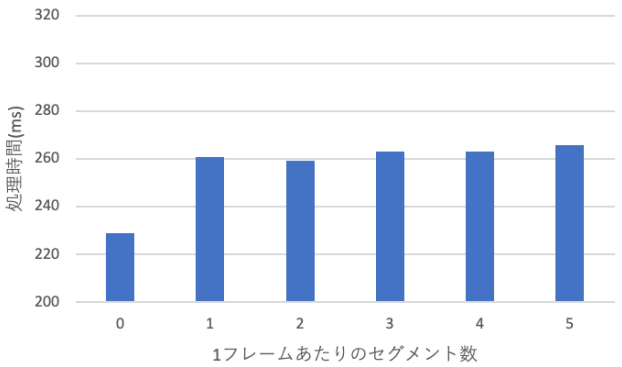


図 10 エッジにおけるセグメント抽出の処理時間

るためである．次に，エッジでデータ選択を行わず，全てのセグメントのデータをクラウドに送信する場合にエッジにおける処理時間を図 10 に示す．図 10 も図 9 と同様に横軸は 1 フレームに対して含まれるセグメントの数を，縦軸にはセグメント毎に抽出に要した時間の平均を示している．セグメント数が 0 の場合の処理時間は 229 ミリ秒，セグメント数が 1 以上の場合はセグメント数にかかわらず，260 ミリ秒程度となっている．これらの結果から，エッジでセグメントの選択を行うために 20 ミリ秒から 40 ミリ秒程度要することがわかる．また，評価に用いたデータでは 935 個のセグメントが検出され，そのうち，候補セグメントは 304 個であった．935 個全てのセグメントに対し，PointNet による属性推定に要する時間は 3.711 秒であった一方，選択された 306 個のセグメントに対する時間は 1.914 秒であった．このことから，データを選択することによりクラウドにおける処理時間が 1.797 秒短くなることがわかる．また，エッジにおける選択を行わない場合の合計処理時間は 148.916 秒である一方，エッジであらかじめ候補セグメントの選択を行う場合の合計処理時間は 159.607 秒であり，エッジにおける 1 フレームあたりの処理時間は 253 ミリ秒から 271 ミリ秒に増加している．そのため，クラウドとエッジ双方に要する合計処理時間は，セグメントを選択しない場合は 152.627 秒，セグメントを選択する場合は 161.521 秒となり，1 フレームあたりの処理時間はそれぞれ

表 3 属性推定を行うセグメントのデータ容量

データ	容量 (バイト)
全てのセグメント	17,380,939
選択されたセグメント	6,092,634

表 4 検出されたセグメントに対するベビーカー利用者属性の推定精度

データ	適合率 (%)	再現率 (%)	f 値 (%)
全てのセグメント	38.3	42.7	40.4
選択されたセグメント	81.8	66.7	73.5

260 ミリ秒と 275 ミリ秒相当となっており，エッジでセグメントを選択するか否かに関わらず，1 秒あたり 3 フレーム程度のデータを処理することができる．利用する三次元深度センサの計測範囲は，奥行き 4m 程度，水平方向 4.6m 程度となっており，時速 4km で歩く場合は，この計測範囲を 4 秒程度で通過することとなる．リアルタイムに歩行者の属性を認識する際には，この 4 秒間に取得される 10 フレーム程度のデータのいずれかに対し，属性を推定することとなり，実世界において十分に機能するものと考えている．

5.1.2 データ量の評価

表 3 に示すように，エッジで検出された全てのセグメントのデータ容量の合計は 17.4MB で，エッジで候補セグメントであると選択されたセグメントのみのデータ容量の合計は 6.1MB である．エッジでセグメントの選択を行うことにより，推定するためにクラウドに送信する必要のあるデータ容量が選択により 3 分の 1 程度まで減少できていることがわかる．

5.1.3 分類精度の評価

エッジで検出された全てのセグメントを PointNet による属性推定，およびエッジで候補セグメントであると選択されたセグメントのみを PointNet による属性推定で，そのデータがベビーカー利用者であるかの評価結果を表 4 に示す．また，それぞれに対しての推定結果の属性の内訳を図 11 と図 12 に混合行列で示す．

表 4 に示すように，検出されたセグメント全てに対して推定を行う場合よりも，エッジで簡易的な判定を行った後に推定を行う場合の方が，適合率，再現率，f 値に対して良い結果が得られた．この結果から，エッジによる選択によりベビーカー利用者属性ではないデータが削除され，ベビーカー利用者属性のデータが多く残されたことがわかる．

適合率が向上した理由として，形状がベビーカー利用者に近いデータについても 1 次近似の係数が大きく異なる場合，そのデータが選択されないことが挙げられる．選択されなかった例として，身長程度の高さの台車を押す人が挙げられる．この場合は，形状はベビーカー利用者に近いが台車の高さが身長程度あるため，1 次近似の係数が小さく

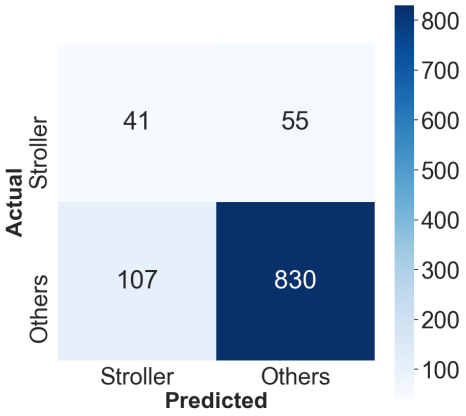


図 11　検出した全てのセグメントに対する属性推定結果の混同行列

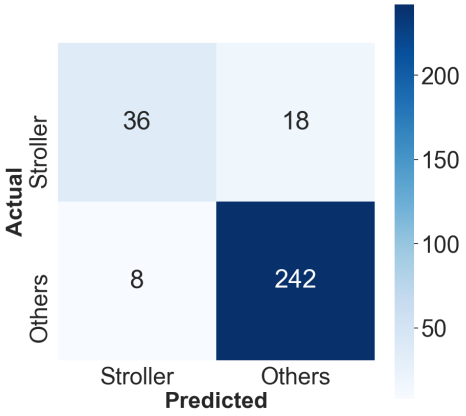


図 12　選択されたセグメントに対する属性推定結果の混同行列

なり，その結果候補セグメントには選択されない．このようにベビーカー利用者属性に形状が似ているが，1 次近似の係数が異なるものが選択されないため，適合率が向上したと考えられる．また，再現率が向上した理由として，図 13 のように取得したデータが画角の端であるなどで欠損のあるデータでは 1 次近似の係数が小さくなるため，属性推定で誤判定を起こしやすいデータが選択されにくいことが挙げられる．

5.1.4　属性推定モデルの精度評価

エッジによる処理とは別に，クラウドにおける属性推定に用いるモデル自体の精度評価を行った．このモデルではベビーカー利用者の他に，歩行者, 子供を抱える人, その他属性の判定を行うことができる．属性推定モデルの学習データに用いた三次元深度データの数は，ベビーカー利用者は 199 個，子供を抱える人は 81 個，歩行者は 200 個，その他の属性は 250 個である．また，テストデータに用いた三次元深度データの数は，ベビーカー利用者は 82 個，子供を抱える人は 20 個, 歩行者は 86 個，その他の属性は 100 個である．

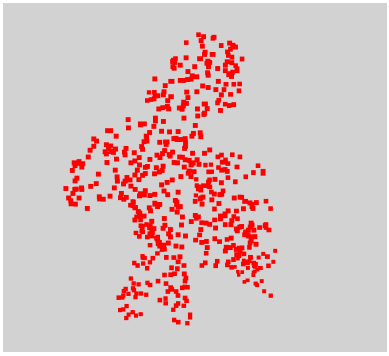


図 13　ベビーカーの半分のデータが欠損したセグメント

表 5　属性の推定結果

属性名	精度
ベビーカー利用者	95.1%
子供を抱える人	55.0%
歩行者	73.3%
その他	99.0%
全体の精度	80.6%

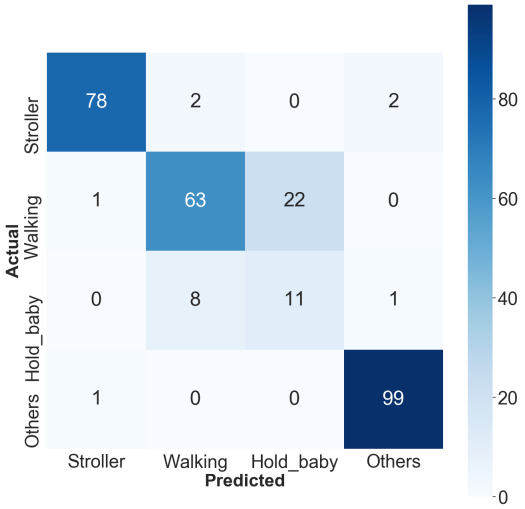


図 14　属性推定結果の混同行列

属性推定の評価を行った結果を表 5 に，図 14 に属性推定結果の混同行列を示す． 各属性の精度は，各三次元深度データに対して四つの属性のいずれであるかの確率を導出し，最も確率の高い属性をその三次元深度データの属性であるとした時の正答率である．属性推定の精度がベビーカー利用者については 95.1 %，子供を抱える人については 55.0 %，歩行者については 73.3 %，その他については 99.0 %であった．また，全体の精度は 80.6 %であった．この結果から，ベビーカー利用者や歩行者のような，形状の大きく異なる属性を高精度で推定できる一方，歩行者と子

供を抱える人のように形状が近い属性に対して誤推定しやすい傾向であることがわかった．

6. おわりに

本研究では，三次元深度データからエッジデバイスでベビーカー利用者の可能性のあるセグメントを抽出し，そのセグメントをクラウドで推定することで，移動制約者の属性を判定する手法を考案した．提案手法では，三次元深度センサーから三次元深度データを取得し，そのデータをエッジデバイスで背景差分とクラスタリングすることにより，移動物体のみの三次元深度データを抽出する．その後，セグメント毎にベビーカー利用者であるか簡易判定を行い，ベビーカー利用者と思われるセグメントのみをクラウドに送信し，クラウドサーバーに配置した PointNet により最終的な属性推定を行う．このエッジデバイスの処理には 271ms, クラウド側での PointNet 判定を含めた処理には 275ms 要することを確認した．また，ベビーカー利用者判定における適合率は 0.818，再現率は 0.667，f 値は 0.735 となり，ベビーカー利用者，歩行者，子供を抱える人，その他の 4 属性の推定を行った結果，80.6%の精度で推定できることを確認した．さらに，エッジでクラウドに送信するデータを選択することにより，データ通信量を約三分の一に抑制することができたことを確認した．

今後の課題として，Raspberry Pi などの小型計算機をエッジコンピュータとし，提案手法を実装することが挙げられる．また，ベビーカー利用者に限らない，多様な移動制約者の把握が挙げられる．歩行を補助する道具の一つとして杖が挙げられるが，そのような道具が必ずしも三次元深度センサーで計測されるものではないため，道具を使っている姿勢や動作を把握することにより，移動制約者の存在を把握する手法についても検討をすすめていく．

参考文献

[1] ユニバーサルデザイン実践の手引き 参考資料編 移動制約者の定 義と配慮事項. https://www.cgr.mlit.go.jp/universal/pdf/01_s01.pdf, 2003. [Online; accessed 2. Feb. 2020].

[2] 災害を受けやすい日本の国土. http://www.bousai.go.jp/kaigirep/hakusho/h18/bousai2006/html/honmon/hm01010101.htm, 2017. [Online; accessed 23. Jan. 2020].

[3] 国土交通省. 高齢者、障害者等の移動等の円滑化の促進に関する法律. http://www.mlit.go.jp/common/001285785.pdf [Online; accessed 8. Feb. 2020].

[4] Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. *arXiv*, 2018.

[5] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[6] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time Human Pose Recognition in Parts from Single Depth Images. In *Computer Vision and Pattern Recognition 2011*, pp. 1297–1304, 2011.

[7] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard. 3-D Mapping with an RGB-D Camera. *IEEE Transactions on Robotics*, Vol. 30, No. 1, pp. 177–187, 2014.

[8] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85, 2017.

[9] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3D Shapenets: A Deep Representation for Volumetric Shapes. In *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, 2015.

[10] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pp. 5099–5108. Curran Associates, Inc., 2017.

[11] C. R. Qi, O. Litany, K. He, and L. Guibas. Deep Hough Voting for 3D Object Detection in Point Clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9276–9285, 2019.

[12] Bo Yang, Jianan Wang, Ronald Clark, Qingyong Hu, Sen Wang, Andrew Markham, and Niki Trigoni. Learning Object Bounding Boxes for 3D Instance Segmentation on Point Clouds. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pp. 6740–6749. Curran Associates, Inc., 2019.

[13] Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[14] Martin Simony, Stefan Milzy, Karl Amendey, and Horst-Michael Gross. Complex-YOLO: An Euler-Region-Proposal for Real-time 3D Object Detection on Point Clouds. In *The European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[15] Z. Zerong, Y. Tao, L. Hao, G. Kaiwen, D. Qionghai, F. Lu, and L. Yebin. HybridFusion: Real-Time Performance Capture Using a Single Depth Sensor and Sparse IMUs. In *Proceedings of The European Conference on Computer Vision*, 2018.

[16] Manuel Marín-Jiménez, Francisco Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 3D Human Pose Estimation from Depth Maps Using a Deep Combination of Poses. *Journal of Visual Communication and Image Representation*, Vol. 55, , 07 2018.

[17] 3D Point Cloud -Supervisely-. https://supervise.ly/lidar-3d-cloud/. [Online; accessed 8. Feb. 2020].

[18] Inc. Occipital. Structure Core - Depth refined. https://structure.io/structure-core [Online; accessed 8. Feb. 2020].