

住居内環境での LiDAR・マイクアレイ統合による移動音源の追跡

Tracking a Moving Sound Source in Indoor-Environment with LiDAR and Microphone-Array

伊福和己 公文誠*
Kazuki Ifuku Makoto Kumon

熊本大学
Kumamoto University

Abstract: 本論文では LiDAR によるポイントクラウドとマイクロホンアレイからの音到来方向情報を統合し、環境中の移動音源位置推定を考える。住居等の場合、環境の多くを占める建物部分が平面でモデル化できることを利用して、形状が未知な環境中の物体を取り出し、音源を認識する方法を提案する。分離された音源候補については、確率的移動モデルと観測の対応づけを利用して、移動音源の追跡も実現する。実環境で行った提案法の評価実験の結果も合わせて示す。

1 はじめに

音環境認識において、音源の位置を正確に推定することは重要で、例えばロボットが対話相手を検出する場合や、災害現場で要救助者を発見しようとする時、また侵入者検知など、この技術が必要となる状況は多々存在する。このような例では、対象以外にも音源が存在することが一般的で、さらに音源自体の発する音が間欠的なことから、音情報のみで対象を継続的・安定して検知、追跡することは容易ではない。特に住環境のような複雑な環境では、壁面での反射や回折など環境との相互作用によって正確な音到来方向の推定 [1, 2] が難しく、マイクロホン（以下マイク）やマイクアレイに加えてカメラ等の異なるセンサを相補的に用いることが効果的である。

カメラ動画像とマイクアレイ情報では音到来方向が得られるため、これらの情報が空間の同一方向を指すことを利用して統合することが考えられる [3]。この場合、対象の奥行き情報は得られないため、例えば複数の音源が交差するような場合、その前後を判断するには複数のカメラや深度情報を用いる [4]、複数のマイクアレイを用いる [5]、カメラとマイクアレイを異なる位置に配置する [6] など、適当な方法で音源位置を推定する必要がある。

対象の奥行き情報を得るために良く用いられるセンサに Laser Imaging Detection and Ranging (以下 LiDAR) があり、多数の計測点からなるポイントクラウドと呼ば

れる点群 (以下点群) で周辺環境を表現する。特に移動する LiDAR の観測情報とオドメトリ情報に Simultaneous Localization and Mapping (以下 SLAM) [7, 8] 手法を援用した環境地図作成は良く研究されており、自動運転技術として実用化の取り組みがなされる段階にある。音環境地図の作成にも展開があり、田邊ら [9] が LiDAR で得られた環境情報である点群と、高精度なマイクアレイによる音到来方向から音源位置を推測する方法を提案している。また、Even ら [10] は LiDAR に基づく SLAM で得られた環境地図に対して、遅延和ビームフォーマで推定した音到来方向情報を統合して音源情報を表す方法を示した。これらはいずれも静的音源を対象としたもので、点群 (あるいはこれを変換したボクセル) と音源の対応を絶対位置を用いて対応づけしている。一方、SLAM で得られる環境地図は静的なものに限られるため、移動物体の検出は別に対応する必要がある。自動運転や屋内作業のようにコンテキストが明確な場合、検出される対象について車、歩行者や自転車などを仮定することは現実的で、点群の中からこれらの物体をセグメンテーションする手法が提案されている (例えば [11, 12])。あるいは移動物体が音を発するという仮定に立脚して、音源付近を取り除いて静的環境の三次元再構成を行った後、移動物体を改めて環境情報に加えるアプローチも提案されている [13]。しかし、日常空間での音源を考える場合、対象は多岐に渡るため、これらを事前に学習しておくことは難しく、また移動物体が常に音を発しているとは限らないので、静的な背景情報から対象物体の候補を分離するのは簡単ではない。特に LiDAR の点群は画像情報に

*連絡先: 熊本大学
〒 860-8555 熊本市中央区黒髪 2-39-1
E-mail: kumon@gpo.kumamoto-u.ac.jp

比べて疎な情報で、対象への遠近によっても疎密が変わることなどから、対象の詳細な形状が得にくいことも考慮する必要がある。

本研究でも環境中に移動音源が存在する場合に LiDAR とマイクアレイで認識することを考えるが、ここでは対象音源ではなく周囲環境を住環境のような一定の性質を仮定し、この特性を用いて点群中から対象を分離することに取り組む。

本論文の構成は次の通りである。次節でマイクアレイによる音源認識についてを概略し、第3節で想定する環境についての条件と、LiDARでの点群情報から対象の候補を抽出する方法、さらに音源の追跡手法について説明する。その後、提案法の妥当性について実験での検証を行い（第4節）、最後に第5節でまとめる。

2 マイクアレイとLiDARによる音環境地図

マイクアレイでの音源方向推定情報は、特に室内のような反射など不確かさのある環境では不正確となるため、複数の観測を統合して推定することが重要である。ここでは、先行研究 (Evenら [10]) を元にマイクアレイと LiDAR を用いて環境中の音源位置を推定し環境地図を構成する方法を概説する。

2.1 音源位置推定

Multiple signal classification (MUSIC) [14] 法などマイクアレイ処理によって音源方向を推定する手法が提案されているが、これらはマイクアレイから見た相対的な方向情報が得られるものの、その奥行き情報を得られない。

本研究では LiDAR を用いて環境地図の形状情報が得られるため、環境地図に音源方向を投射するアプローチを採るものとする。また音源方向の推定には、耐雑音性が高く、複数音源からの到来方向にも対応可能とされる MUSIC 法を用いることとする。具体的には MUSIC 法によって得られた MUSIC スペクトルが十分に大きな値を示した時、これに対応する方向を音到来方向とする方法を考える。

環境地図情報を \mathcal{M} と表し、推定された音源方向 θ が得られたとき、マイクアレイから音源方向に向けた線分が環境地図との交点 x を求めることが出来る。

$$x = x(\theta, \mathcal{M}) \quad (1)$$

ただし、 θ には一定の不確かさがあるため、推定された音源位置 x にも不確かさがあるため、この不確かさを考慮する必要があるが、 x が音源であるかどうかは確率的に表現されることになる。

2.2 環境地図中の音源確率

時刻 t において推定された音源方向 $\theta(t)$ について MUSIC スペクトルが $P_M(\theta(t))$ とする。この観測に対して、(1) で与えられる x に音源があるという信念を表す尤度 $L(x(\theta(t)), \mathcal{M})$ を以下で与える。なお、以降、誤解のない範囲で \mathcal{M} は省略する。

$$L(x(\theta(t)), P_M(\theta(t))) = p_{min} + \frac{p_{max} - p_{min}}{1 + e^{-\frac{P_M(\theta(t)) - T_a}{\alpha}}} \quad (2)$$

ここで p_{min}, p_{max} は尤度の範囲を定めるパラメータであり、 α, T_a は尤度分布を定める定数である。

今、 \mathcal{M} の点 A が音源であるという事象を E_A 、 A が (1) で求まる $x(\theta(t))$ と一致する、あるいは近傍にあるという事象を $S_{A,\theta(t)}$ と表す。 $S_{A,\theta(t)}$ の下で A が音源である確率 $P(E_A|S_{A,\theta(t)})$ は Bayes の定理と、 $P(\bar{E}_A) = 1 - P(E_A)$ の関係より、

$$P(E_A|S_{A,\theta(t)}) = \left[\frac{1 + (1 - P(S_{A,\theta(t)}|E_A))(1 - P(E_A))}{P(S_{A,\theta(t)}|E_A)P(E_A)} \right]^{-1} \quad (3)$$

となる。対数オッズ関数

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (4)$$

を用いれば (3) は以下のように変形できる。

$$\text{logit}(P(E_A|S_{A,\theta(t)})) = \text{logit}(P(S_{A,\theta(t)}|E_A)) + \text{logit}(P(E_A)) \quad (5)$$

また (5) を繰り返し用いることで、このような観測を N 回得た時 ($S_{A,\theta_1:\theta_N}$ などと書く) 、

$$\text{logit}(P(E_A|S_{A,\theta_1:\theta_N})) = \text{logit}(P(S_{A,\theta_N}|E_A)) + \text{logit}(P(E_A|S_{A,\theta_1:\theta_{N-1}})) \quad (6)$$

のように漸進的に計算することが出来る。実際には観測を得た際の確率 $P(S_{A,\theta})$ を (2) の尤度に置き換えて、

$$\text{logit}(P(E_A|S_{A,\theta_1:\theta_N})) = \text{logit}(P(E_A|S_{A,\theta_1:\theta_{N-1}})) + \text{logit}(L(x(\theta_N), P_M(\theta_N))) \quad (7)$$

とする。これを推定された音源方向情報全てについて計算することで音環境地図を構成する。なお、初期確率は適宜与えるが、 $P(E_x) = \frac{1}{2} (\forall x \in \mathcal{M})$ とすることが多い。

3 住居内環境での音源検出

3.1 住居内環境の音源

本研究で考える住居内環境についてまず説明する。

住居は人の生活に供する空間であるが、改めてその構造を考えると、外界から守るために壁面・屋根等の境界を与える建物部分があり、その内部に机や戸棚等の什器などの物体と人の移動・滞在できる空間がある。Manhattan 仮説 [15] が指摘するように人工物の環境構造には平面や平行、直交などの特性があるため、本研究でも住居の特に建物部分については平面によって構成されたと考える。実際、住居内で観測をする場合、床、天井、壁面によって全方向を覆われる状況になるため、LiDAR の観測する点群の多くの部分は平面でモデル化できる特徴がある。また、音源となりうる家具等は壁際に存在することが多く、例えば壁掛け時計などを考えれば、壁と物体を点群中から適切に区別することは工夫が必要である。一方、住居内の音源は数個程度が孤立していて自由空間部分を移動するものが含まれる可能性があり、建物そのものが音を発することは少ない。移動音源が壁面近くに滞在する可能性は注意が必要である。

以上より、本研究では LiDAR で得た対象の環境の点群には、いくつかの平面を構成する多数の点と、それ以外の要素を成す独立したクラスタが存在すると考え、これらの独立したクラスタが物体を表しているというモデルに立脚して各物体を追跡する手法を提案する (図 1)。また、この物体には静的物体と移動物体双方を含むことに注意されたい。

3.2 静的平面の検出

前述の環境に対する仮定から、壁などの建物については平面構造で面積が大きいため観測点数が多く、観測中における点群位置の変化は微小という性質がある。このような環境を構成する点群は、大きなクラスタを構成するため、比較的特定が容易であることから環境を構成する点群の内、平面状に分布する点群を検出し、近似平面に分類された点群を建物環境として扱う。

近似平面の検出には Random Sample Consensus [16] (RANSAC) を用いて、クラスタサイズの逆数と平面モデルとの二乗距離和から成る評価関数に基づいて法線と原点からの距離に対応する平面パラメータを推定する。さらに、平面として検出された点群を取り除いた後、残る点群に対して繰り返し同手法を適用して環境の平面点群を複数検出する。この手続きは平面として検出されたクラスタサイズが検出を実行した入力点数の一定割合未満となるまで続ける。以下 N_p 個の平面が検出されたと考え、これらの平面パラメータを $\phi_p = \{\phi_{p,1}, \dots, \phi_{p,N_p}\}$ などと書く。

また、対象とする平面は静的と考えているため、パラメータ ϕ_p は観測時刻によらず一定となる筈である。時刻列 t_1, \dots, t_T におけるスキャンデータを前述の方法

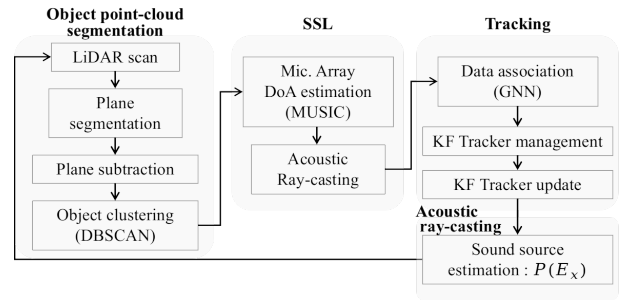


図 1: Flow of the proposed method

で処理して得たパラメータの集合 $\{\phi_p(t_1), \dots, \phi_p(t_T)\}$ をパラメータ空間で Density Based Spatial Clustering of Applications with Noise (DBSCAN) [17] によってクラスタリングし、一定以上のサイズを持つクラスタの重心を静的平面として推定し、これらに所属する点群によって静的環境情報 \mathcal{M}_S を定義する。この平面群に対応する点群を各時刻での LiDAR からの点群から取り除き、環境中の静的物体と移動物体のクラスタを示す点群を得る。

3.3 環境内物体の検出・追跡

前小節の方法で得た環境中物体の点群を物体毎に認識するため、点群をクラスタに分類しこれらのクラスタの動きを推定する。なお、以下では簡単のため 2 次元平面内の記述とする。

3.3.1 個別物体の識別

環境中の多くを占める建物由来の点群を前小節の方法で推定し取り除くことで比較的少数から成る物体に対する点群 (以下物体点群と呼ぶ) が得られる。この物体点群に含まれる各物体毎の点群クラスタは互いに非連結となっていると期待されることから、物体点群に対して DBSCAN [17] を適用して、各物体毎の点群クラスタ (以下物体クラスタ) を得る。このように時刻 t において k_t 個に分類された物体クラスタを $\{C_{t,1}, \dots, C_{t,k_t}\}$ と表すこととする。

3.3.2 個別物体の追跡

住居環境での移動は、その空間的制約から速さや動きは制限されていることから、移動物体には適当な運動モデルを想定することは現実的である。本研究ではシンプルな例として等速直線運動を導入する。

時刻 t で l 個の追跡対象が得られており、その m 番目の物体点群の中心位置を $(x_{t,m}, y_{t,m})$ とする。状態

を $\mathbf{x}_{t,m} = (x_{t,m}, \dot{x}_{t,m}, y_{t,m}, \dot{y}_{t,m})^T$ とし、等速直線運動の仮定の下で状態方程式は以下のように書ける。

$$\mathbf{x}_{t,m} = \mathbf{F}\mathbf{x}_{t-1,m} + \mathbf{G}\Delta\mathbf{x}_{t-1,m} \quad (8)$$

$$\mathbf{F} = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \frac{\tau^2}{2} & 0 \\ \tau & 0 \\ 0 & \frac{\tau^2}{2} \\ 0 & \tau \end{bmatrix}$$

ここで τ はステップ間のサンプリング周期で、 $\Delta\mathbf{x}_{t,m}$ は平均 $\boldsymbol{\mu}_{x,t-1,m}$ 、共分散 $\mathbf{Q}_{t-1,m}$ の正規加速度外乱とする。

一方、クラスタの重心が物体位置として観測される。時刻 t の物体点群のうち n 番目のクラスタ $C_{t,n}$ が物体 m のものとすれば、 $C_{t,n}$ の重心座標 $\mathbf{z}_{t,n}$ によって観測方程式は以下の式で表される。

$$\mathbf{z}_{t,n} = \mathbf{H}\mathbf{x}_{t,m} + \Delta\mathbf{z}_{t,n} \quad (9)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

ここで $\Delta\mathbf{z}_{t,n}$ は平均 $\boldsymbol{\mu}_{z,t,n}$ 、共分散 $\mathbf{R}_{t,n}$ の正規外乱とする。

上記モデルから、以下に示すカルマンフィルタを用いて動きの予測を行う。

1. 予測ステップ

$$\bar{\mathbf{x}}_{t/t-1,m} = \mathbf{F}\mathbf{x}_{t-1,m} + \boldsymbol{\mu}_{x,t-1,m} \quad (10)$$

$$\bar{\mathbf{P}}_{t/t-1,m} = \mathbf{F}\mathbf{P}_{t-1,m}\mathbf{F}^T + \mathbf{G}\mathbf{Q}_{t-1,m}\mathbf{G}^T \quad (11)$$

2. 更新ステップ

$$\mathbf{y}_{t,mn} = \mathbf{z}_{t,n} - (\mathbf{H}\bar{\mathbf{x}}_{t/t-1,m} + \boldsymbol{\mu}_{z,t,n}) \quad (12)$$

$$\mathbf{S}_{t/t-1,m} = \mathbf{H}\bar{\mathbf{P}}_{t/t-1,m}\mathbf{H}^T + \mathbf{R}_{t,n} \quad (13)$$

$$\mathbf{K}_{t,m} = \bar{\mathbf{P}}_{t/t-1,m}\mathbf{H}^T\mathbf{S}_{t/t-1,m}^{-1} \quad (14)$$

$$\mathbf{x}_{t/t,m} = \bar{\mathbf{x}}_{t/t-1,m} + \mathbf{K}_{t,m}\mathbf{y}_{t,mn} \quad (15)$$

$$\mathbf{P}_{t,m} = \bar{\mathbf{P}}_{t/t-1,m} - \mathbf{K}_{t,m}\mathbf{H}\bar{\mathbf{P}}_{t/t-1,m} \quad (16)$$

なお、(10) で得られる状態 $\bar{\mathbf{x}}_{t/t-1,m}$ に対して、次節で求める対応関係が得られる場合のみ更新ステップ (12)~(16) を行なうものとする。対応する観測が得られない場合は

$$\mathbf{x}_{t/t,m} = \bar{\mathbf{x}}_{t-1,m}$$

$$\mathbf{P}_{t,m} = \bar{\mathbf{P}}_{t-1,m}$$

とする。

3.3.3 データアソシエーション

実際の状況では、有効領域内に観測が複数の観測を得る場合や、複数の追跡位置から重複して観測される場

合があり、追跡位置と観測値との対応を求めるデータアソシエーションが必要となる。本研究ではこの対応付けに演算の簡便な Global Nearest Neighbor (GNN)[18]を用いる。

観測された物体クラスタ n と推定している物体 m との対応を求めるため、物体クラスタ n の観測位置と物体 m の推定位置の間の距離にマハラノビス距離

$$d_{t,nm}^2 = \mathbf{y}_{t,nm}^T \mathbf{S}_{t/t-1,m}^{-1} \mathbf{y}_{t,nm} \quad (17)$$

を考える。GNN 法では、まず有効領域の閾値 g と、領域外を表す十分に大きな正定数 E を定めて次のコスト関数 λ_{nm} を考える。

$$\lambda_{nm} = \begin{cases} E & d_{nm}^2 > g \\ d_{nm}^2 & d_{nm}^2 \leq g \end{cases} \quad (18)$$

(18) を全ての観測・対象の組合せについてまとめたコスト行列 $\boldsymbol{\Lambda}$ を以下のように定める。

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & & \vdots \\ \lambda_{l1} & \lambda_{l2} & \cdots & \lambda_{lk} \end{bmatrix} \quad (19)$$

物体 m に対応する観測値の番号を $a(m)$ ($m \neq m'$ に対して $a(m) \neq a(m')$) とした時、 $\sum_{m=1}^l \lambda_{m,a(m)}$ を最小とする $a(m)$ が求める対応関係となる。この最小化には Munkres 法 [19] を用いることで効率的に計算できる。

3.3.4 追跡管理

物体の追跡プロセスの経過とともに、新しい追跡対象の出現や、それまでの追跡対象が消失することがあり、それに応じて対応するカルマンフィルタを生成・消去する管理機構が必要である [20]。

(a) 新規追跡物体の生成

得られた観測が現在追跡中のいずれのカルマンフィルタの推定値からもマハラノビスの意味で十分に離れている場合、この観測を追跡する新たなカルマンフィルタが開始される。ただし、観測値が突発的な外乱によるものである可能性を想定し、追跡開始から一定回数 (N_1 と記す) 以内に次の観測が一度でも得られない場合は追跡を中止する。

(b) 追跡の終了

オクルージョンなどで追跡している物体に対する観測が一時的に得られないことがある。観測が得られなくなった場合、予測ステップ (10) (11) を用いて物体の追跡を続けるが、長時間にわたって観測が得られない場合、推定値は信用できないものとなる。そこで十分

な回数（ここでは N_2 回とする）続けて観測の得られない状況が続いた時に当該の物体についての追跡を終了する。

以上で求められた時刻 t での各物体の点群によって物体環境情報 $M_{O,t}$ を与える。これから、環境情報は $M = \{M_S, \{M_{O,t_1}, \dots, M_{O,t_T}\}\}$ と得られる。また $M_{O,t}$ は速度の推定も含んでいるため閾値処理等で移動物体と静止物体を区別する。

3.4 音源情報との統合

Ray-cast 法は、指定した場所から透明な光線を放ち、光線と交わる物体から情報を取得する方法であり、ここでは簡単のため、球との交点を ray-cast 法で求め、音源位置を推定する。

時刻 t における音源方向 θ_t が得られたとする。物体クラス $C_{t,j} \in M_{O,t}$ の点群 $\{c_{t,j}^1, \dots, c_{t,j}^i, \dots\}$ の全ての点それぞれについて、その点を中心とする半径 r の球を考え、マイクアレイから音源方向 θ_t に向けた半直線との交点の有無を求める。交点が存在すれば、物体クラス $C_{t,j}$ から音が到来したと判断する (図 2)。複数のクラスと交点を持つ場合、これらいずれか、あるいはそのうちのいくつか、または全てから音が到来している可能性があるため、全てを音源候補と考えて、前述の音源確率の計算を行う。

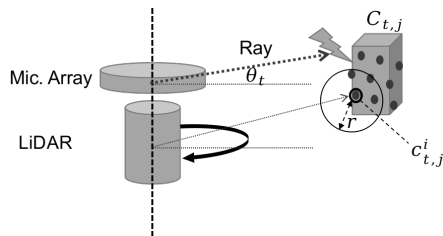


図 2: Cluster based Acoustic Ray-casting

4 検証

提案法の有効性を確認するため、居室内 (図 3) での検証実験を実施した。提案法の実装において、音源定位にはロボット聴覚用オープンソースソフトウェア HARK[21] を用い、ロボットミドルウェア ROS[22] にて LiDAR のスキャンデータと同期して収録した。また、クラスタリング等の点群処理には Open3D[23] を利用した。

本研究で用いたマイクロホンアレイと測定装置を図 4 に示す。実験装置を構成する LiDAR は垂直視野 30 度、水平視野 360 度の範囲の点群が取得可能な、Velodyne 社製の VLP-16-LITE を使用し、円上に 8 つのマイク

が配置されたマイクアレイ (System Infrontier 社) と音響処理ユニット RASP-ZX (System Infrontier 社) を用いて音信号を収録した。マイクロホンアレイと LiDAR は三脚上で軸を同一となるよう固定した。

住環境において想定した静止物体としては室内の壁面、床、天井の他、机、ホワイトボード、洗面台、PC モニタ等を考え、追跡物体のオクルードを検証するために装置中心から x 方向に -0.35m , y 方向に -1.02m 離れた地点に幅 0.71m の壁を設定した。また、環境中を移動する音源として、人が継続的に発声を行いながら装置の周りを楕円状に移動した。この移動はスタート地点からマーカーに沿って時計回りに人間の歩行速度で周回するものとした。得られたスキャン点群ならびに MUSIC スペクトルの時間発展を図 5 と 6 に示す。

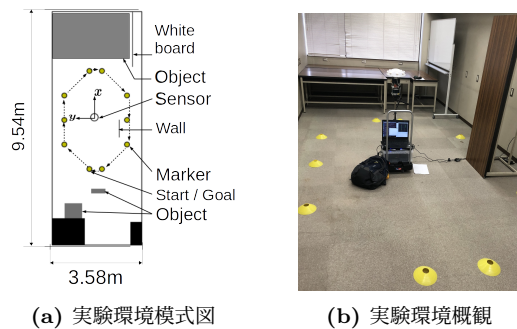


図 3: 実験環境

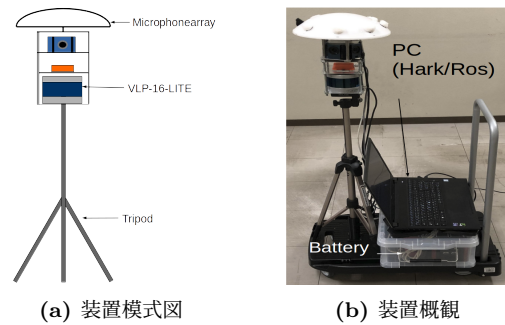


図 4: 実験装置

今回のシステムでは、LiDAR のスキャン周期 $\tau = 0.1[\text{sec}]$ を基準に同期して処理しており、提案法のカルマンフィルタに用いたパラメータ値は $N_1 = 5$, $N_2 = 80$, $Q = 0.1I$, $R = 0.01I$ とした。音源確率の推定では $p_{min} = 0.1$, $p_{max} = 0.9$, また予備実験から $\alpha = 40$, $T_a = 20$ とした。

まず環境から建物部分とそれ以外の住居内物体の区別の様子を図 7 に例示する。クラスタ毎に色分けしており、壁面、床、天井などの建物に対し、仮説の通り室内の人、机等が非連結のクラスタとして分けられていることが分かる。

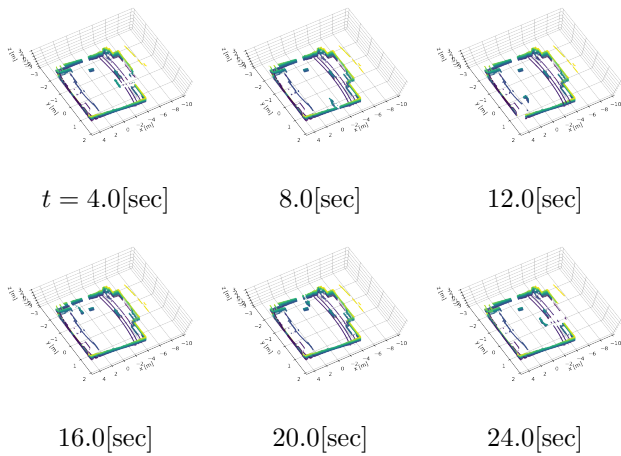


図 5: スキャンされた点群

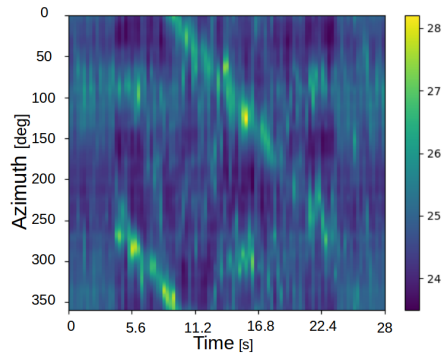


図 6: MUSIC スペクトログラム

次に各物体クラスターの追跡結果を図 8 に示す。図中黒い点は平面検出によって建物として認識された点群を成し、赤い点が物体とされた点群を示している。物体中の黒点がカルマンフィルタの平均、その周辺の等確率楕円で推定された不確かさを示しており、0.1[m/sec]以上の速さが推定された物体については、矢印でその速度ベクトルを表している。室内中央を楕円経路に沿った移動音源を良く検出・追跡している。20[sec]付近から図右下付近にも移動音源を推定しているが、これはオクルージョンに伴って、静止物体がいくつかの小領域に分かれたため、いくつかの小領域の重心がオクルードしている移動音源の動きに合わせて誤って推定されたものである。この推定は共分散の大きな不確かな情報で、適当な閾値処理などで区別することが可能と考えられる。

提案法はクラスタリングとカルマンフィルタの追跡によって、移動音源の点群についても一貫した音源確率の推定を行うものである。これを確認するため、各物体で推定された音源確率を図 9 に示す。提案法 (図 9(a)) では移動音源に対して高い確率を与える一方、室内に

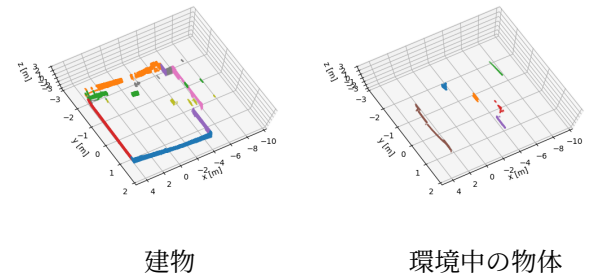


図 7: クラスタ分類結果 ($t = 20.0[\text{sec}]$)

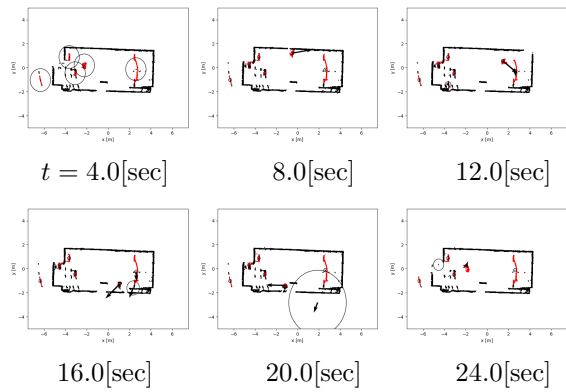


図 8: 移動音源・物体追跡

あった音を発さない物体については低い確率となった。比較のため、クラスタ追跡を行わなかった場合の結果を図 9(b) に示す。MUSIC スペクトルに強弱があったため、閾値処理によって音源の検出出来なかった区間では音源推定確率が低くなっており、また非音源との区別も難しい。このことを定量的に評価するため、音源確率を元に音源の存在判定した場合の適合率 V_P と再現率 V_R を図 10 に示す。なお、音源確率の分布範囲が両者で異なるため、正規化して閾値処理するものとし、試行中の検出すべき延べ音源数を Num_S とし、正しく検出した音源の総数を Num_T 、音源として検出した総数を Num_D とした時 適合率は $V_P = \frac{\text{Num}_T}{\text{Num}_D}$ 、再現率は $V_R = \frac{\text{Num}_T}{\text{Num}_S}$ で与えられる。図より適切な閾値において

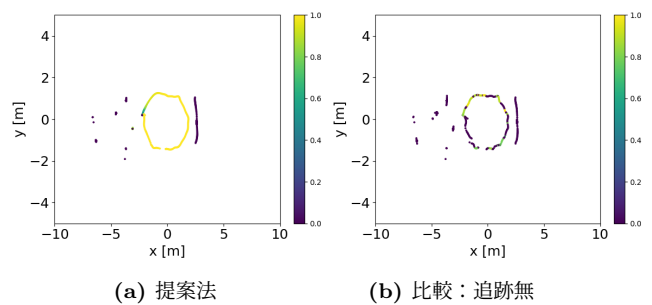


図 9: 推定された音源確率

追跡を行わない場合に比べ提案法の方が性能の良いこと、特に再現率において優位であることが示された。

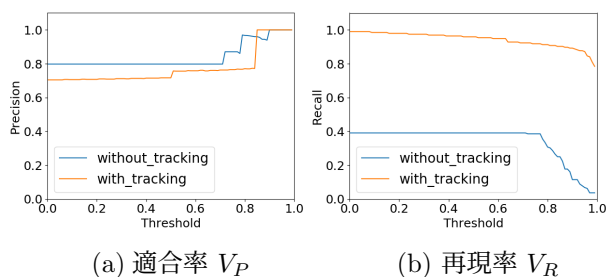


図 10: 性能評価

5 おわりに

本研究では、住居内環境でLiDARを用いて点群を得る時、建物部分を適当な仮定の下で取り除いて室内物体と区別し、物体と音到来方向情報を統合して音源を認識する手法を提案した。複数の物体が室内にある場合でも、対象を区別できることを実験で示した。また、対象が移動物体であっても音源である確率を推定できることも確認し、提案法の有効性が示された。オクルージョンが発生する場合、物体クラスタの認識に影響することがあり、今後改善が必要である。

謝辞

本研究はJSPS 科研費 19H00750 の助成をうけた。

参考文献

- [1] G. Narang et al., “Auditory-aware Navigation for Mobile Robots based on Reflection-robust Sound Source Localization and Visual SLAM,” Proc. of the 2014 IEEE Int. Conf. on Sys., Man, and Cyb. (SMC2014), 2014, pp.4079-4084.
- [2] K. Takami et al., “Estimation of a nonvisible field-of-view mobile target incorporating optical and acoustic sensors,” *Autonom. Robot.*, 40(2), 2016, pp. 343-359.
- [3] Y. Kokusho and M. Kumon, “Sound Source Tracking by Incorporating Target Motion Estimated by Visual Trackers,” Proc. of Int. Sym. on Sys. Integ. (SII2020), 2020, pp. 652-657.
- [4] 鈴木啓他, “環境音情報と画像情報を用いた物体検出による音ラベル付きセグメントの生成” 日本ロボット学会 学術講演会予稿集, 2020, 1D3-02.
- [5] K. Sekiguchi et al., “Online Simultaneous Localization and Mapping of Multiple Sound Sources and Asynchronous Microphone Arrays,” Proc. of Int. Conf. on Intel. Robot. Sys. (IROS2016), 2016, pp. 1973-1979.
- [6] 公文誠, 鷲崎海他, “繰り返しベイズ推定を用いた視聴覚統合による話者位置推定,” 計測自動制御学会システムインテグレーション部門講演会, 2018, 3B3-13.
- [7] J.J. Leonard and H.F. Durrant-whyte, “Simultaneous Map Building and Localization for an Autonomous Mobile Robot,” *Intelligent Robots and Systems (Int. Workshop at IROS91)*, 1991, pp. 1442-1447.
- [8] 友納正裕, SLAM 入門, オーム社, 2018.
- [9] 田邊亮他, “マイクロホンアレイと3次元LIDARを用いた確率的音源地図作成,” ロボティクスメカトロニクス学術講演会, 2016, 1P1-09b4.
- [10] J. Even et al., “Probabilistic 3D Mapping of Sound-Emitting Structures Based on Acoustic Ray Casting,” *IEEE Tr. on Robot.*, 33(2), 2017, pp. 333-345.
- [11] RB. Rusu et al., “Towards 3D Point Cloud based Object Maps for Household Environments,” *Proc. IEEE Int. Conf. on Robot. and Autom.*, 2007, pp. 927-941.
- [12] M. Lehtomaki et al., “Object Classification and Recognition From Mobile Laser Scanning Point Clouds in a Road Environment,” *IEEE Tr. on Geo.Sci. and Rem. Sens.*, 54(2), 2016, pp. 1226-1239.
- [13] T. Konno et al., “Audio-Visual 3D Reconstruction Framework for Dynamic Scenes,” *Proc. of IEEE/SICE Int. Symp. on Sys. Integ. (SII 2020)*, 2020, pp. 802-807.
- [14] R.Schmidt, “Multiple emitter location and signal parameter estimation” *IEEE Tr. on Ant. and Prop.*, 34(3), 1986, pp. 276-280.
- [15] J. M. Coughlan and A. L. Yuille. “The Manhattan World Assumption: Regularities in Scene Statistics which Enable Bayesian Inference,” *Proc. of Int. Conf. on Neural Info. Process. Sys.*, 2000, USA, 809-815.
- [16] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Comm. ACM.* 24(6), 1981, 381-395.
- [17] M. Ester et al., “A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proc. Int. Conf. on Know. Discov. and Data Min.*, 1996, pp. 226-231
- [18] P. Konstantinova et al., “A Study of a Target Tracking Algorithm using Global Nearest Neighbor Approach,” *Proc. CompSysTech*, 2003, pp. 290-295.
- [19] J. Munkres, “Algorithms for the Assignment and Transportation Problems,” *J. Soc. Indust. Appl. Math.*, 5(1), 1957, pp. 32-38.
- [20] M. Wakabayashi et al., “Multiple Sound Source Position Estimation by Drone Audition Based on Data Association Between Sound Source Localization and Identification,” *IEEE Robot. and Autom. Lett.*, 5(2), 2020, pp. 782-789.
- [21] K. Nakadai et al., “Development, Deployment and Applications of Robot Audition Open Source Software HARK,” *J. of Robot. and Mech.*, 29(1), 2017, pp. 16-25.
- [22] M. Quigley et al., “ROS: an open-source Robot Operating System,” *ICRA workshop on Open Source Software 3 (3.2): 5*, 2009.
- [23] Q.Y. Zhou et al., “Open3D: A Modern Library for 3D Data Processing,” *arXiv:1801.09847*, 2018.