

科学館での自律移動ロボット実証実験 —ロボットは人混みで音を聞き分けられるか!?!— Demonstration Experiment in a Science museum — Can the robot understand the sound in the crowd!?! —

佐々木洋子^{1*} 坂東宜昭¹ 金崎朝子¹
Yoko Sasaki¹ Yoshiaki Bando¹ Asako Kanazaki¹

¹ 国立研究開発法人産業技術総合研究所

¹ National Institute of Advanced Industrial Science and Technology

Abstract: The paper introduces our demonstration experiment of an autonomous mobile robot in a science museum. The goal of the project is developing the autonomous mobile robot that works in the crowded situation. The experiment was started on September 2016 and is ongoing project. As the first step, we start understanding “crowded” situation by observing pedestrians, and clarifying what the simple navigation can and cannot do without environment model. It confirmed that the mobile robot can observe pedestrians’ behavior while moving and the simple navigation can be effective when there are not so many people, but cannot work well in the crowds at narrow aisle. Based on the results, we start the second step of the demonstration experiment on July 2018. The goal is acquisition of “auditory function” and “navigation” that works in the crowded situation. We just started developing the algorithms and recording the sensor data to solve the problems by machine learning approach.

1 はじめに

通勤時間帯の駅構内、週末の繁華街など、人でも移動に気を遣うような雑踏環境で動作する自律移動ロボットの実現を目指し、1) 雑踏環境を自在に動くためのロボット動作生成、2) 雑踏環境で状況を理解するためのリアルタイム音環境認識、の2つを課題として、科学館での長期実証実験を行っている。本稿では、研究室での基礎的な機能開発と実用化の中間にあたる、科学館で実施中の実証実験の取り組みについて紹介する。

画像処理分野では変化の激しいアクション映画にも耐える画像認識技術 [1] 等がある一方、冒頭の2つの課題「動作生成」と「聴覚」は、いずれも街中の雑踏環境下で十分に動作する技術はまだ確立されていない。移動ロボットの基本機能である動作生成、周囲の情景理解に不可欠な聴覚機能と、どちらも雑踏環境での自律移動ロボットに欠かせない機能である。そこで本研究では、雑踏環境の実フィールドとして位置づけられる科学館での長期実証実験を通して、機械学習アプローチで実験と学習・手法改善を繰り返しながら、モデル化困難な雑踏環境を克服する取り組みを開始した。

*連絡先：国立研究開発法人 産業技術総合研究所
〒135-0064 東京都江東区青海 2-3-26
E-mail: y-sasaki@aist.go.jp

2 科学館での実証実験



図 1: An experimental snapshot in Miraikan

人の身近で活動する自律移動ロボットの開発は、研究室での基礎開発だけでは完成せず、多くの人が活動する中での実証が不可欠である。本節では、科学館で進めている自律移動ロボット実証実験の概要をまとめる。

2.1 研究のゴール

本研究では、雑踏環境下で自在に動作可能な自律移動ロボットの実現を最終ゴールとし、2016年9月より、不特定多数の人が往来する科学館での実証実験を行っている(図1)。実験の第一段階として、2017年度末までの1年4ヶ月間は、イベント名「ロボットは自分で

人混みを抜けられるか!?’と題して、混雑環境での移動機能に着目した研究を行った。ここではまず、(1) 移動ロボットによる周囲の歩行者観測、(2) 既存の自律移動機能の混雑下での課題抽出、の2つを研究テーマとして、実験を開始した。この第一段階では、環境の観測および、既存手法の有効性と限界の確認が、主な課題となっている。また後半では、前半に蓄積した実験データを用いて、動作生成の学習に取り組んだ。

第二段階として2018年度7月からは「ロボットは人混みで音を聞き分けられるか!?’と題して、雑踏環境下での音環境認識をメインテーマに実験を開始した。前年度までの実験第一段階で得られた知見から、第二段階では「聴覚」と「動作生成」を重点課題としてあげた。雑踏環境では人に囲まれ視界も限られることから、周囲の情景理解には聴覚機能が重要な役割を果たす。一方、無数に音が存在し刻々と変化する動的環境では、未だ有効な音環境認識技術は存在しない。動作生成についても同様に、第一段階で実証した従来法では雑踏環境の走破性に限界がある。いずれも動的に変動する未知の環境に対する頑健性が不可欠であり、実証実験第二段階では実験と学習/手法改善を繰り返しながらモデル化が困難な雑踏環境の克服を目指す。なお本論文の執筆時点では、実証実験第二段階についてはまだ研究の初期段階であるため、第一段階で得られた成果の概要と、および第二段階で開始した「聴覚」についての取り組みを紹介する。

2.2 実証実験環境

不特定多数の人が往来する雑踏環境として、実証実験を進めている日本科学未来館について説明する。日本科学未来館は、様々な国籍、年代の人が日に数千人訪れる科学館で、実験は展示フロア3階で実施している。会場のフロアマップを図2に図示する。ここは5階とともに常設展示フロアとなっており、各階の面積はおよそ3000[m²]である。図に示す青色部分は来館者の見学エリアで、特に順路等はなくエリア内を自由に回る展示形式となっている。緑矢印は主な人の流れを示す。図中央の入り口および、右側に5階へ続くスロープの2カ所の出入り口がある。

赤字のマーカおよび矢印は、あらかじめ設定したウェイポイント(通過点)を示している。実験ではこのウェイポイントを巡回するようにロボットが自律移動する。さらに赤字の番号位置には歩行者観測の精度評価用にレーザーセンサを設置した。三脚上、高さ1mの位置にセンサを設置し、その下に小型PCとモバイルバッテリーを入れており、三脚ユニット単体で動作する。

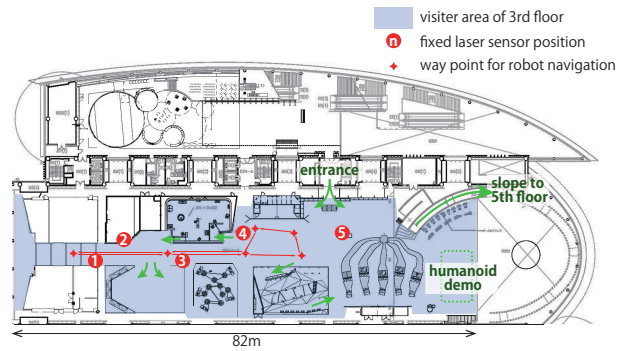


図2: Floor plan of the experiment field : Miraikan

3 自律移動ロボット | Peacock

本節では、実証実験に使用している自律移動ロボット“Peacock”の仕様を説明する。

3.1 要求仕様

本実証実験に使用するロボットには、以下の条件で動作することが求められる。

- 面積 3000[m²] の段差のない屋内環境
- 順路等はなく来館者が自由に往来する環境
- 開館から閉館まで7時間の連続動作

多数の人が往来する中で、人や展示物に衝突することなく動き続けることを前提とし、車輪型ロボットに周囲の環境認識と障害物回避機能を搭載した自律移動ロボット“Peacock”を開発した。

3.2 ハードウェアの構成

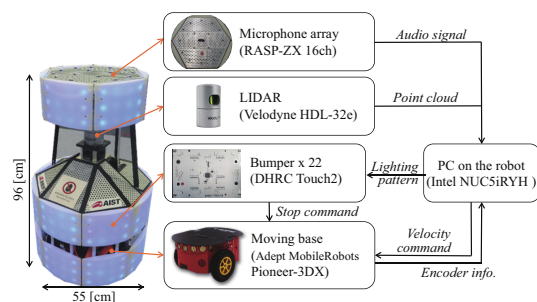


図3: Hardware configuration of Peacock

図3にPeacockのハードウェア構成を示す。Peacockは最大幅55cm、高さ96cmの円筒形で、ベースのロボット台車にはPioneer-3DXを使用している。外装は22枚

のバンパー基板からなり、非常停止およびLEDによる外観装飾の機能を持つ。各基板の四隅にあるタクトスイッチがPioneer-3DXのバンパー回路に接続されており、進行方向側スイッチのいずれかが押されるとモータが停止する設計となっている。また各基板の3×4個のLEDはPCに接続され、ソフトウェアから制御することができる。周囲環境の認識用に、LIDAR(Velodyne HDL-32e)および、16chマイクロホンアレイ(RASP-ZX)を搭載している。

以上は全てロボットに搭載した小型PC(Intel NUC kit)に接続されており、全ての計算はロボット内で行う。PCの電源は、電動自転車用のリチウムイオン電池(26v, 8.1Ah)を使用し、約4時間動作する。バッテリーソケットを2つ搭載することで、PCを停止することなくバッテリー交換が可能である。PC以外の機器には、Pioneer-3DX内のリン酸鉄リチウムイオンバッテリー(12v, 9Ah × 4個)から給電する。7時間以上の連続動作が可能であり、同時に転倒防止のためPeacockの重心を下げる役割も担っている。

3.3 ソフトウェアの構成

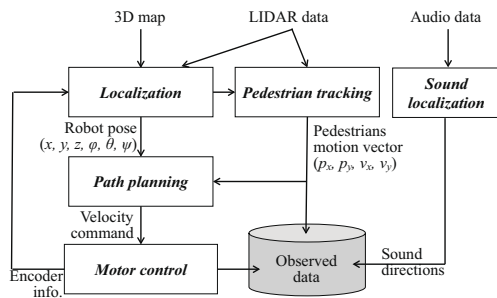


図 4: Software configuration of Peacock

ソフトウェアの全体構成を図4に図示する。まずLIDARから得られる3D点群を用いて、あらかじめ作成した環境の3次元点群地図に対するPeacockの6自由度位置・姿勢を推定する。同時にロボット周囲の歩行者追跡と障害物検出を行う。これらの結果を経路計画モジュールで2D占有格子地図に反映し、ゴールまでの最短経路探索を行う。算出された経路に沿う速度指令をモータに出力し、モータのエンコーダ値を自己位置推定モジュールに戻す。一連の計算をLIDARの観測周期と同じ10[Hz]で繰り返すことで、多数の人がいる環境下での自律移動を実現する。同時に、直接自律移動には使用していないが、マイクロホンアレイで音源定位を行い、Peacock上部に取り付けたリボン状のLEDで検出した音の到来方向を表示する。

また「聴覚」と「動作生成」の学習のため、すべてのセンサデータおよび各モジュールの計算結果は時刻情報と合わせて保存する。

3.3.1 自己位置推定

事前に作成した会場の3次元点群地図を事前知識として利用し、地図に対する6自由度位置・姿勢をリアルタイムに推定する[2]。本手法は、内界センサにIMU、外界センサにLIDARを使用し、モンテカルロ位置姿勢推定(Monte-Carlo Localization, MCL)に基づき、与えられた3次元形状地図に対する6次元の位置姿勢を確率分布として推定する手法である。ロボットの動作モデルが不要で、センサユニット単体で動作するため、手持ちのような任意方向への複雑な動きをはじめ、様々な移動体に対応可能である。Peacockでは平面内を車輪移動するため、IMUの代わりに車輪エンコーダの値を位置・姿勢推定の初期値に使用する。

3.3.2 歩行者追跡

LIDARを用いた歩行者観測手法として、角度ベース複数仮説を用いた複数移動体追跡[3]を行う。本手法は、水平に設置された単一スキャンラインのLIDARを対象にした移動体追跡手法で、周囲の多数の歩行者に対して、これまでの移動軌跡、現在位置、移動速度ベクトルを出力する。本手法では、まず前処理として、環境地図と一時刻前のスキャンデータをもとに点群を前景と背景に分割し、前景点群を移動体候補として抽出する。次に、水平スキャンから見た歩行者を楕円体と仮定することで、パーティクルフィルタを用いて歩行者の外形を追跡する。このとき、追跡対象とスキャン点群の角度範囲との対応付けに複数の仮説を持たせ、最尤仮説を求めることで近接した移動体の切れ目を時系列に判断しながら複数の移動体を追跡することができる。前処理段階で点群のセグメンテーションを確定せず、切れ目も含めて追跡することで、近接した複数の歩行者に強い手法となっている。なお、本稿では歩行者のみを対象としたが、本手法そのものは、それぞれ外形を定義することで複数種類の移動体を同時追跡することが可能である。

3.3.3 自律移動

観測した周囲の障害物および歩行者の情報を2次元占有格子地図に反映し、まずA*アルゴリズムに基づく最短経路探索により、次のウェイポイントまでの経路計画を行う。次に、算出した経路に沿って動くように、Dynamic Window Approach(DWA)を用いて、モータへの速度指令を計算する。経路計画の実装には、ROSのナビゲーションパッケージを利用した。

図5に占有格子地図の作成手順を図示する。本研究では経路計画でロボットを点として扱うため、以下の占有格子地図作成時には、障害物範囲をロボットの幅

分増幅させる。3D 地図を変換した静的レイヤと、各時刻での観測に基づく動的レイヤの大きく 2 層からなる。まず自己位置推定に用いている 3 次元点群地図 a) から、ロボットの高さに相当する点群を抽出し、2 次元平面に投影し、占有格子地図に変換する b)。本研究では、床面を基準として高さ 0.3 から 1.2[m] の点群を抽出し、グリッドサイズは 5[cm] とした。動的レイヤには、各時刻で観測した LIDAR 点群と歩行者追跡の結果を用いる。自己位置推定結果に基づき、各時刻で得られる LIDAR のデータ (c) ベージュ点群) と検出した歩行者 (c) 青ベクトル) を地図座標系に投影する。移動体の位置は変化するため、ロボット近傍のデータのみ、静的レイヤと同様にロボットの高さ分を占有格子地図に投影する d)。このとき歩行者については、速度ベクトルに応じて仮想的に障害物範囲を拡大する。b) と d) を重ねた占有格子地図 e) を経路計画に用いる。

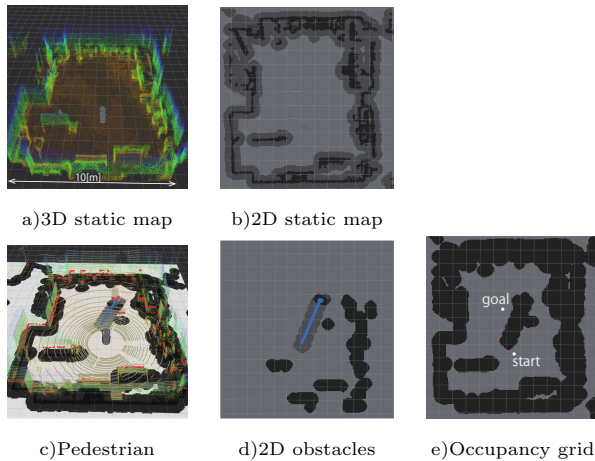


図 5: 2D occupancy grid map for navigation

経路計画では、図 5e) に示す占有格子地図に対し、ロボットの現在位置から次のウェイポイントまでの最短経路を A* アルゴリズムで探索する。得られた経路に対し、DWA でモータへの速度指令 (速度, 角速度) を計算する。DWA の評価関数には、以下の 3 項を用いる。

- 占有格子地図上で障害物と衝突しない
- 経路に沿う (経路線からの距離)
- ゴールを向く (ゴール方向と進行方向のなす角)

また障害物のそばでは安全のために、最近傍障害物までの距離 d に応じて移動速度 V を減速させる。

$$V = d/L_1 \times V_m (d < L_1),$$

V_m は最大速度, L_1 は距離の閾値を表し、本実証実験では、 $V_m = 0.5[\text{m/s}]$, $L_1 = 1.0[\text{m}]$ とした。

3.3.4 音源定位

自律移動には直接利用しないが、来館者に見せるデモンストレーションの一環として、Multiple Signal Classification (MUSIC) による音源定位を行う。実験では、MUSIC で検出した音の到来方向は Peacock 上部に取り付けた LED および、説明用の大型ディスプレイに表示する。本研究の目的は、実環境音を大規模に収集し、雑踏環境の音環境認識を実現することであるが、科学館での展示イベントとして実験を進めるために、データ収集と合わせて、既存技術で可能なことと限界を伝えることを目的としてデモンストレーションを行っている。また音に反応して Peacock が光るため、多くの人に話しかけてもらえるという副次的な効果も得られる。

4 実験事例紹介

2016 年 9 月より月に 2 日ペースで実証実験を継続している。実験を通して収集しているデータ、および、これまでの研究事例をいくつか紹介する。

4.1 移動ロボットによる歩行者観測

Peacock の特長のひとつとして、動きながら周囲の歩行者観測が可能であることが挙げられる。実証実験の第一段階では、まず Peacock が移動しながら観測した歩行者データの解析を行った [4]。ここでは歩行者観測の解析結果について、概要を紹介する。

2016 年 9 月から 2017 年 6 月までの 10 ヶ月間、延べ 140 時間強のデータを用いて、次の 2 つの課題について検証を行った。

- Peacock が移動しながら歩行者を観測可能か
- ロボット周囲の人の動きは通常と異なるか

移動中の Peacock から観測した歩行者軌跡と、実験エリア全体をカバーするよう 5 カ所に設置した評価用レーザセンサの観測を比較した結果、移動しながら設置センサと同様に歩行者を観測可能であることを定量的に示した。ただし視点が異なるため、検出結果は完全に一致はしない。一方で、特に混雑した状況では、人と人の間隔が狭く隠れや人数変化も大きいため、ロボット/設置センサ双方とも、LIDAR 視野内の歩行者のうち観測できたのは 6 割程度になる時間帯もあった。

ロボット周囲の人の動作については、実験エリアを 3m グリッドに区切り、単位時間あたりの通過人数と平均歩行速度を、ロボット/設置センサそれぞれで算出し比較した。なおロボットのデータについては常に実験エリア全体を観測できないため、LIDAR データから各セルの観測時間を算出し補正した。ロボット/設置セン

サそれぞれの結果を比較すると、通過人数の分布では、ロボットの観測結果が全体的に人数が多いものの、人数の大小分布は同様の傾向が得られた。この結果は、ロボット周囲は通常より人が多いこと、常に全体を観測していない移動ロボットからでも、人が頻繁に通る場所といった人流の分布傾向を観測可能であることを示している。また平均歩行速度の分布については、比較的広い場所ではロボット/設置センサとも同様の分布が得られたが、特に狭い通路状の場所でロボットが観測した結果が顕著に速度が遅くなった。この結果は、狭い場所でロボットが立ち往生するため、人の流れを妨げているためと考えられる。

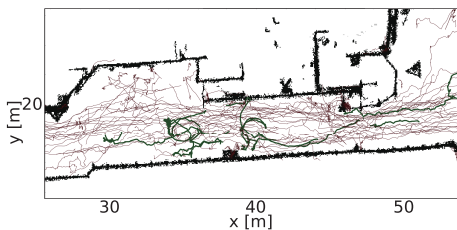


図 6: Pedestrian trajectories observed from Peacock

以上の結果を示す一例として、Peacock から観測した歩行者軌跡の例を図 6 に示す。10 分間に観測した歩行者軌跡を占有格子地図上に描いた。この中で、Peacock からの距離 1[m] 以内に連続 5[sec] 以上とどまっていた軌跡を緑線に表示した。比較的長い Peacock に付いていく軌跡や、人をよけようと向きを変える Peacock を「通せんぼ」する円弧状の軌跡が確認できる。実験ではこのような行動が頻繁に見られ、この結果が Peacock 周囲に人が多く、狭い場所で歩行速度が遅い現象につながったと考えられる。

4.2 CNN による動作生成

前節で述べた最短経路探索に基づく自律移動では、

- 人や物に衝突することなく終日自律移動が可能
- 比較的人が少ない状況では十分な動作

であることを確認できた。一方、混雑する時間帯になると、人に囲まれゴールまでの経路がないため立ち往生する、動かないためさらに混雑する、という状況がたびたび発生した。そこで、ロボット周囲の状況から進行方向を決定する手法、GOSELO(Goal-Directed Obstacle and Self-Location Map for Robot Navigation) [5] による自律移動を行った。GOSELO では、ゴールまでの経路探索をする代わりに、これまでの経路、ゴール方向および周囲の占有格子地図から変換した GOSELO マップを入力に、CNN で次の進行方向を決定する。入

力する GOSELO マップは、占有格子地図 3 チャンネル、これまでの経路履歴 3 チャンネルからなる計 6 チャンネルの画像である。まずゴール方向が上向きになるよう占有格子地図を回転させ、ゴールと現在位置の中間地点を画像中心にする。次に地図の 3 チャンネルについては、ゴールと現在位置との距離 L を基準長さとし、一辺が $(L+4)$, $4 \times L$, $8 \times L$ の範囲が異なる 3 枚を占有格子地図からクロップする。経路履歴の 3 チャンネルについては、地図と同範囲について、各セルの通過回数を整数値で入れる。CNN のネットワーク構造は AlexNet [6] を用いる。上記の 6 チャンネルの画像を入力とし、最後の全結合層は 10 度刻み、36 方位の進行方向を出力する。

本研究では、収集したデータから、各時刻でのゴール位置と占有格子地図および過去の自己位置推定結果を GOSELO マップに変換し、 A^* による経路探索結果を 36 方位に離散化した進行方向を教師として、ネットワークを学習した。教師が A^* 探索結果であるため、人とうまくすれ違うといった動きを獲得するわけではないが、設定したウェイポイントを終日巡回することができ、またゴールまでの経路がないような混雑した状況であっても動き続けることができた。

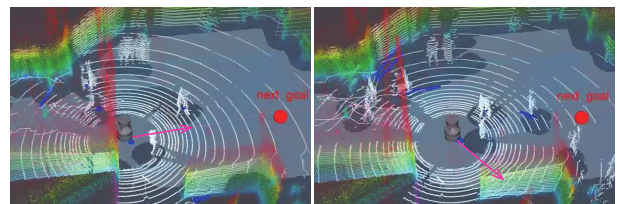


図 7: Snapshots of GOSELO navigation

2017 年 8 月に実施した GOSELO による自律移動の様子を図 7 に示す。ROS の可視化ツール“rviz”のスナップショットである。色つき点群で 3 次元点群地図、床面上に占有格子地図および次のウェイポイント(赤マーカー)を表示している。さらに推定した自己位置に合わせて、ロボットモデル、LIDAR 入力(白色点群)、歩行者追跡結果(青ベクトルと白線)、および GOSELO 出力(マゼンタベクトル)を描画している。歩行者の位置に合わせて、Peacock の進行方向が変わっていることが確認できる。このほか、アルゴリズムおよび実験結果の詳細については、[5] を参照されたい。

4.3 音源分離の教師なし深層ベイズ学習

2018 年 7 月より開始した実証実験の第二段階では、雑踏環境下での音環境認識に取り組んでいる。まだデータ収集を開始した段階であり、実際のデータを使った結果は得られていないが、現在学習を進めている教師なし音源分離手法 [7] と、シミュレーションデータを用いた結果を紹介する。

本研究で対象とする「雑踏環境」は、無数に音が存在し時々刻々と変化するため、従来法のように音源数を仮定したり、音源定位→分離→識別と順序処理するアプローチでは対応できない。そこで動的な実環境に耐える手法として、音源分離を空間情報に基づいて教師なし学習する手法を提案した [7]。近年、高い性能を達成している深層学習による音源分離には、あらゆる音源の多数の教師データを要するという課題がある。一方、これまで広く研究されている空間情報に基づく多チャンネル分離は少ない事前知識で動作する利点があるが、同一方向の音源を分離できない原理上の限界がある。提案法では両者を活用し、深層学習による単チャンネル音源分離を空間情報を用いて教師なし学習するアプローチをとる。提案法は多数の多チャンネル録音があれば学習でき、音源分離の実行時には、単チャンネルマイクロホンでも動作し、同一方向の音源分離も期待できる。ここでは、空間モデルの一手法である Latent Dirichlet Allocation (LDA) [8] を用いた音源分離を扱う。本モデルは潜在変数として音源分離を行う時間周波数マスクを持ち、マスク推定に基づく深層学習を用いた音源分離と相性が良い。

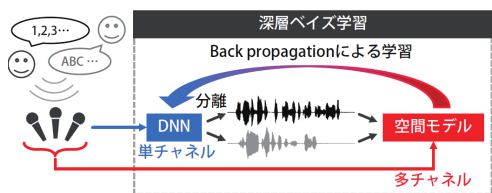


図 8: Unsupervised sound separation approach

提案法の基本性能を評価するため、WSJ-0 英語読み上げ音声から音源数=2 ずつ混合したシミュレーションデータで音源分離の学習を行った。生成データは、8チャンネル円形マイクロホンアレイに無響室インパルス応答を畳み込み、音源方向はランダムに与えた。提案法と同じ確率モデルを、空間情報のみ用いる変分ベイズ推論したものと比較した結果、音源が別方向にある場合にやや性能が劣るものの、同一方向の場合でも別方向と同精度で分離できることを確認した。学習やネットワーク構成に改善の余地が示唆されたものの、提案する空間情報に基づき教師なし学習するアプローチの有効性を示すことができた。今後、ノンパラメトリックベイズに基づく音源数の同時推定・拡散性雑音項の導入といったモデル改善を行うとともに、実証実験で収集したデータによる学習を進めていく。

5 むすび

本稿では、雑踏環境下で自在に動作する自律移動ロボットの実現を目指し、科学館で進めている長期実証実験の取り組みを紹介した。2016年9月より開始した

本実証実験では、まず「雑踏環境」を理解し課題を抽出するため、歩行者観測と基礎的な自律移動の有効性検証を行った。この結果、動き回るロボットから歩行者の分布傾向を観測できること、ロボット周囲では人の動きが変化することを定量的に示すことができた。また、最短経路探索に基づく自律移動は人が少ない状況では十分に機能する一方、雑踏環境での限界も明らかになった。これらの結果を踏まえ、機械学習アプローチに基づく「聴覚」と「動作生成」の獲得を目標に、2018年度7月から実証実験の第二段階を開始した。今後、データを収集しながら学習と手法改善を繰り返すことで、複雑な雑踏環境に対応することを目指す。

謝辞

本実証実験は日本科学未来館にて実施中である。実証実験イベントの企画・運営を進めて頂いた伊達雄亮氏、伊藤健太郎氏、酒井尚子氏をはじめ、日本科学未来館スタッフの皆様には感謝する。また本研究の一部は、科研費 研究活動スタート支援 No. 18H06490 の支援を受けた。

参考文献

- [1] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [2] Akihito Ohsato, Yoko Sasaki, and Hiroshi Mizoguchi. Real-time 6dof localization for a mobile robot using pre-computed 3d laser likelihood field. In *Proceedings of International Conference on Robotics and Biomimetics*, pp. 2359–2364, December 2015.
- [3] Hatao Naotaka and Satoshi Kagami. Cluster-based sjpdafs for classification and tracking of multiple moving objects. In *Proceedings of 9th Conference on Field and Service Robotics*, 2013.
- [4] Yoko Sasaki and Jirou Nitta. Long-term demonstration experiment of autonomous mobile robot in a science museum. In *Proceedings of IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, Ottawa, Canada, October 2017.
- [5] Asako Kanazaki, Jirou Nitta, and Yoko Sasaki. Goselo: Goal-directed obstacle and self-location map for robot navigation using reactive neural networks. *IEEE Robotics and Automation Letters*. *Journal of Robotics and Mechatronics*, Vol. 3, No. 2, pp. 696–703, April 2018.
- [6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097–1105, 2 2012.
- [7] 坂東宜昭, 佐々木洋子. 音源分離のための潜在ディリクレ配分法の教師なし深層ベイズ学習. 第 36 回日本ロボット学会学術講演会講演論文集, pp. 1E3–05, 2018.
- [8] Takuma Ohtsuka, Katsuhiko Ishiguro, Hiroshi Sawada, and Hiroshi G. Okuno. Bayesian nonparametrics for microphone array processing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 2, pp. 493–504, February 2014.