

令和 2 年 6 月 5 日現在

機関番号：17401

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00365

研究課題名（和文）無人航空機による地上物体認識のための視聴覚統合

研究課題名（英文）Audio-Visual Integration to Target Recognition by Drone Audition

研究代表者

公文 誠（Kumon, Makoto）

熊本大学・大学院先端科学研究部（工）・准教授

研究者番号：70332864

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本課題ではドローンから地上音源を検知する上で、ドローン自身の発する騒音が大きく、また対象音源が遠くにあつて、マイクロホンからの信号だけで音源を正確に認識することは難しいことを想定し、カメラとマイクロホンアレイを搭載した無人航空機で安定かつ高精度に対象を認識する技術を目的とした。

間欠的な音情報と連続的だが音源の外見について不確かな動画情報を相補的に統合し、音源の三次元位置推定、複数移動音源の識別・追跡などを実現した。

研究成果の学術的意義や社会的意義

本来音によって特徴づけられる音源について、一定の条件の下で画像という異なるセンサ情報（外見の分からない音源）を通じて認識する出来るようになったことを通じ、様々なセンサ情報を自律的に統合する方向へと展開が可能で意義があると考えている。異常検知や、防犯等、社会一般で必要とされる技術としても利用可能である。

研究成果の概要（英文）：In this study, it is considered to recognize targets on the ground from drones with microphones. The target acoustic signal obtained at the drone is generally significantly distorted by the ego-noise, and, hence, it is difficult to recognize the target only by acoustic signals. This study aims to develop the technology to compensate this difficulty by incorporating visual sensor information.

Acoustic features that contain pauses is fused with visual features that are normally provided sequentially where it is not trivial to associate the visual information with the acoustic target. Based on the developed methods, it is shown that audio-visual integration improves the audio target recognition under noisy situation, and as an example, three-dimensional position estimation of moving plural targets by the drone with microphones was achieved.

研究分野：ロボット工学

キーワード：ロボット聴覚 視聴覚統合 センサフュージョン ドローン聴覚

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

研究計画・開始時において、無人の小型マルチロータヘリコプタ(ドローン)を活用した空撮などの業務が始まっており、より広範な応用が期待される状況にあった。申請者は関連する先行研究において複数のマイクロホンから成るマイクアレイを搭載したドローンを用いて地上音源の位置推定のテーマに取り組んでおり、叫び声のような定常的な大きな音で環境中に孤立した音源であればある程度検出できることを確認しつつある段階にあった。

一方、これらの従来の研究では一定の条件の下でその可能性を示唆することに注力しており、例えば複数の音源が存在したり、これらが移動する場合などより複雑な環境においてマイクやマイクアレイを搭載したドローンを用いて音源位置を精度良く推定するといったより広範な活用には、新しいアプローチの必要性があると考えていた。

提案当時でもカメラを搭載したドローンを用いた空撮は一般的であったため、この映像情報を音源位置情報に活用することが可能であれば、ドローンの応用において自然な展開であり音源位置推定に有効な手段であると想定され研究を始めることとなった。

2. 研究の目的

無人航空機から地上を監視する際、例えば走行する車両の追尾や火災・建築の崩落などで、音の発生源を注視するように、様々な地上の物体の中から興味の対象を検知することは極めて重要である。このような技術の実現を目指し、本課題ではドローンから地上音源を検知することを想定し、カメラとマイクロホンアレイを搭載した無人航空機で安定かつ高精度に対象を認識する技術に資するマルチモーダルな音源認識技術の確立を目的とする。

無人航空機では、自身の発する騒音が大きく、また対象音源が遠くにあることから、マイクロホンからの信号だけで音源を正確に認識することは難しく、一方、カメラだけでは音源を検出することは出来ないため、カメラ・マイクロホン情報からの視聴覚統合の基本技術としてマルチモーダルステレオと呼ぶ音源定位の枠組みを構築することとした。

3. 研究の方法

複数カメラを用いたステレオビジョンによる環境技術を、視覚情報と聴覚情報のように異なるモダリティを包含するよう一般化するマルチモーダルステレオ技術が本研究の基礎をなすが、具体的にはカメラ・マイクロホンアレイによる視聴覚統合を通じて音源の方向・距離(奥行)を同時に推定することを例題に研究を遂行することとした。これを実現するには、異なるセンサ群から得られた対象の特徴点同士の対応点の求め方、およびこれらの異なる種類のセンサ情報をどのように統合するかの二点において基礎理論が不可欠で、2017年から1年程度注力した。実際、ステレオビジョンでは対応点を精度良く求めることが奥行き情報を推定する上で重要であるように、カメラ・マイクロホン双方での検出される方向情報の不確かさ・精度に着目して検討することでこの問題に取り組んでいる。

この対応点検出では、センサ群の特徴量が異なった特性を持つことから技術的困難が生じている一方、音情報には音源が音を発しない時間区間が存在するのが一般的なものに比べ、画像情報は時間的に連続して得られる、というように異なるモダリティ情報を相補的に利用することで、推定性能を向上することも想定される。2018年～19年にかけて、このように適切に特徴の利点を取り扱うアルゴリズムについて検討し、より高性能な情報統合の手法を構築した。このアプローチは異なるセンシング情報を統合するものであるため、より一般的な複数センサによる音源定位手法としてカメラ以外のセンサ情報との統合とした一般化にも取り組んでいる。

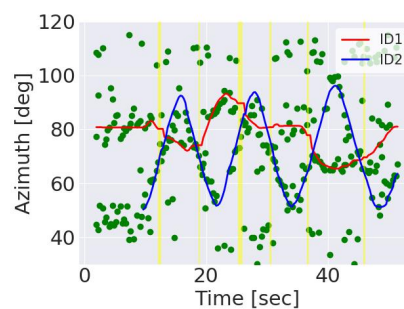
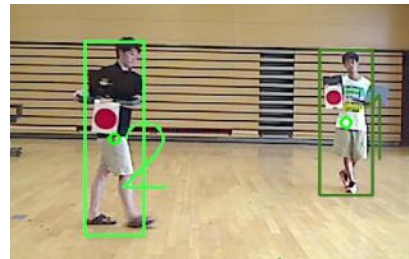
4. 研究成果

4.1. カメラ・マイクロホンからの視聴覚情報統合による音源定位

異なる種類(モダリティ)のセンサからの情報として視覚・聴覚を対象に考え、これらを

統合することで対象音源の位置・方向を推定する技術として検討した。特に、ドローンでの音収録は、ロータなどの騒音により目的音が明瞭に聞こえない・途切れがちになる問題があり、これを視覚によって補うことを一つの応用として想定した。

カメラ・マイクロホンアレイとも、対象音の方向を検出することは出来るものの、その奥行きを正確に得ることは一般には容易ではない。これらのセンサが両眼視の場合と同様に一定の距離（ベースライン）を持って配置されれば、三角測量の要領で対象の3次元位置を推定可能となる。そこで、まず対象音源の外見に一定の事前知識（例えば対象が人間であるなど）を仮定し、離れて設置されたカメラ・マイクロホンアレイの対を考え、ビデオカメラの動画像とマイクロホンアレイのマルチチャンネル音信号から対象(人)への方向情報を繰り返しベイズ推定の結合尤度として扱う方法を考察した。実際に体育館程度の広さの空間で複数の人が歩きまわる場合で検証した所、奥行きを含めた平面位置を得られた(右図)。グリッドベース繰り返しベイズ推定は演算量が多いことが技術的問題となることがあるが、建物内程度の規模であれば計算アルゴリズムを工夫することでノートPCのCPUのみで実時間実行できる実装を行い、また対象を特に話者(人間)に限定することで、効果的に追尾が可能であることを示した。



複数音源定位実験の様子と定位結果

次に対象音がどのような画像上、どのような特徴（外見）を有するかが不明な場合にも適用することを考え、音源定位(方向)情報を利用して動画像より音源領域の検出し、その領域をグリッドベース繰り返しベイズ推定により追跡する手法や、演算の高速なカルマンフィルタによる近似追跡手法を提案した。これらの手法において、複数の音源が互いに交差するような判別の難しい状況にあっても適切に音源追跡が可能であることを実験的にも確認した。特にカルマンフィルタでの実装では、マイクロホンアレイから見た音源方向と想定画像領域の運動の一致を手がかりに、画像中の音源の表現に依らず追尾出来る点、進展があった。このことを利用して、昨今進展の著しい機械学習手法などによって特徴量間の対応を与えるデータ・セットを構築することを期待し、センサ対で収録した試験データから音・画像の組合せを自動的に検出し、音・画像の特徴データの組を記録するシステムも開発した。

これらの研究を通じ、画像中に対象がオクルードされることと、音源信号が雑音等でマスクされる、あるいは発話が中断する(音が途切れる)ことが等価であることから、いずれのモダリティにおいても同じアプローチが可能であることから、以下で示すようにカメラ・マイクに限らない情報統合へと一般化可能であるとの示唆を得た。

4-2. 複数音源の定位

ドローンによる音源定位では、空中を飛来するドローンのマイクロホンで受信する範囲が広く、その対象領域内に複数の音源（非目的の音源も含む）が存在するのが一般的であり、それらの信号はロータ雑音の下で微弱な混合音として収録されることになる。従って、複数の音源を区別して、定位することはドローン聴覚における重要な技術である。ここで、各フレームで得られる定位情報がどの音源のものかを知るためにはさらなる付加的な情報が有用で、複数の情報源（センサ）の統合を目的とする本課題の主題と合致する。

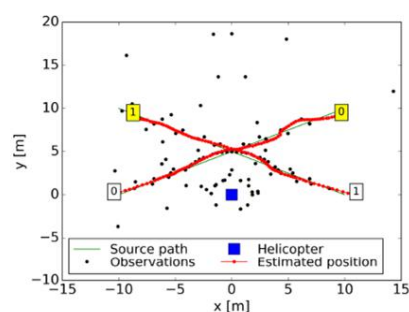
本課題では、視覚におけるオクルージョンと同様に話者同士がマイクロホンアレイから見て同じ方向にある場合に定位が困難になる、いわば音源方向推定におけるオクルージョンの問題があることに着目し、このように音源がほぼ同一方向にある場合を例題として考え、音源定位手法を考案した。従来の音到来方向の推定手法では不確かな結果が

得られるものの、一定の仮定の下で定位に先立って音源を分離するなどして、対象音の周波数情報を用いて定位性能を改善できることを示した。さらに、複数の音がほぼ同じ方向にある場合、これらの音源方向をそれぞれ正確に求めることは難しく、特に環境中にある騒音(ファン騒音など)方向に話者が重なる場合、話者の位置のみを正確に得るには、騒音成分を除外して定位する必要がある。このため、非負値分解を利用して音信号を分離し、その情報を用いて音源方向推定を行う手法を考案し、音源同士が水平方向で5度程度まで接近しても話者方向が推定できるものとなった。

このような音のオクルージョンは、ドローンでの音源定位以外にも、例えば楽器同士の音が重なっている音楽から、個別の楽器の音信号を推定する場合も同様であり、このための手法も考案している。また、環境中の複数音から、各音源のストリームを分離するセグメンテーション手法についても検討した。

4-3. ドローンによる地上音源定位

上述のように、複数の音源が存在する場合において、特に本課題の意義がある。目的であるマルチロータヘリコプタでの実際の活用を想定し、地上に複数の音源がある場合のマルチロータヘリコプタでの収録音による定位についても検証を行った。ドローンで収録される複数の対象の発する混合信号の適切な分離・識別するは音源追跡の重要な要素である。本課題ではロボット聴覚技術を利用した音源分離と bag-of-words のアプローチに基づいた音源の区別方法を提案し、二つの移動音源が交差するような場合であっても、安定して音源追跡が可能であることを示した。この方法では、音源方向情報に加え、音信号の音響的な周波数領域の特徴量を機械学習の手法によって定量化することで、付加的なモダリティ情報を生成する構成になっている。さらに、複数の音源が存在する時、音源位置の推定情報の不確かさが非ガウスとなることを明らかにし、このことを踏まえたドローンの飛行経路を計画するアクティブ音源探査手法も開発した。また、収録するマイクロホンアレイが複数存在する(複数のドローンが飛行する)ことで複数センサを活用することで三次元音源定位を行う研究も実施した。



対象としたマイクアレイ搭載ドローンと移動二音源追尾結果(数値例)

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 中臺 一博	4. 巻 74-7
2. 論文標題 音響信号処理の変遷と最先端 ロボット聴覚の視点	5. 発行年 2018年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 394-400
掲載論文のDOI（デジタルオブジェクト識別子） 10.20697/jasj.74.7_394	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kenzo Nonami, Kotaro Hoshiba, Kazuhiro Nakadai, Makoto Kumon, Hiroshi G Okuno, Yasutada Tanabe, Koichi Yonezawa, Hiroshi Tokutake, Satoshi Suzuki, Kohei Yamaguchi, Shigeru Sunada, Toshiyuki Nakata, Ryusuke Noda, Hao Liu	4. 巻 128
2. 論文標題 Recent R&D Technologies and Future Prospective of Flying Robot in Tough Robotics Challenge	5. 発行年 2019年
3. 雑誌名 Disaster Robotics, Springer	6. 最初と最後の頁 77-142
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-05321-5_3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kotaro Hoshiba, Kai Washizaki, Mizuho Wakabayashi, Takahiro Ishiki, Makoto Kumon, Yoshiaki Bando, Daniel Gabriel, Kazuhiro Nakadai, and Hiroshi G. Okuno	4. 巻 17-11
2. 論文標題 Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments	5. 発行年 2017年
3. 雑誌名 Sensors	6. 最初と最後の頁 2935
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/s17112535	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 中臺 一博	4. 巻 71-5
2. 論文標題 オープンソースコミュニティに貢献すること	5. 発行年 2017年
3. 雑誌名 映像情報メディア学会誌	6. 最初と最後の頁 647-653
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 中臺 一博	4. 巻 62-2
2. 論文標題 マイクロホンアレイを用いた音源定位・分離ソフトウェア入門	5. 発行年 2018年
3. 雑誌名 システム/制御/情報	6. 最初と最後の頁 42-49
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計21件 (うち招待講演 3件 / うち国際学会 2件)

1. 発表者名 國生悠斗, 公文誠
2. 発表標題 屋外での移動音源追跡における動画像と音信号の統合の検討
3. 学会等名 第36回日本ロボット学会学術講演会
4. 発表年 2018年

1. 発表者名 Makoto Kumon, Kai Washizaki, Kazuhiro Nakadai
2. 発表標題 Close Sound Source Localization Inproprating Semi-Supervised Variational Bayesian NMF
3. 学会等名 Int. Symposium on System Integration (SII2019) (国際学会)
4. 発表年 2019年

1. 発表者名 公文誠, 鷲崎海, Tomonari Furukawa
2. 発表標題 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
3. 学会等名 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
4. 発表年 2018年

1. 発表者名 山田 泰基, ガブリエル ダニエル, 糸山 克寿, 西田 健次, 中臺 一博
2. 発表標題 複数の移動マイクロホンアレイによる移動音源の3次元定位
3. 学会等名 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
4. 発表年 2018年

1. 発表者名 山田 泰基, Daniel Gabriel, 糸山 克寿, 西田 健次, 中臺 一博
2. 発表標題 複数のマイクロホンアレイを搭載した複数のUAVによる移動音源の三次元追跡
3. 学会等名 情報処理学会第81回全国大会
4. 発表年 2019年

1. 発表者名 中臺一博
2. 発表標題 ロボット聴覚研究とその展開 - 災害時の迅速な要救助者発見 に向けたドローン聴覚技術開発に至るまで
3. 学会等名 蔵前技術士会 (招待講演)
4. 発表年 2018年

1. 発表者名 中臺一博
2. 発表標題 ドローン聴覚～災害地での迅速な要救助者の捜索を目指して～
3. 学会等名 ImPACTシンポジウム (招待講演)
4. 発表年 2018年

1. 発表者名 干場功太郎, 中臺一博, 公文誠, 奥乃博
2. 発表標題 アクティブ周波数レンジフィルタを用いた雑音にロバストな音源定位手法の提案
3. 学会等名 人工知能学会 第49回AIチャレンジ研究会
4. 発表年 2017年

1. 発表者名 若林瑞保, 公文誠
2. 発表標題 マイクロホンアレイを有するマルチロータヘリコプタを用いた地上の複数音源の位置推定について
3. 学会等名 人工知能学会 第49回AIチャレンジ研究会
4. 発表年 2017年

1. 発表者名 干場功太郎, 中臺一博, 公文誠, 奥乃博
2. 発表標題 UAV搭載マイクロホンアレイを用いた組み込みシステムによる音源探査性能の評価
3. 学会等名 第35回日本ロボット学会学術講演会
4. 発表年 2017年

1. 発表者名 鷺崎海, 公文誠, 大塚琢馬, 奥乃博, 干場功太郎, 中臺一博
2. 発表標題 マルチロータヘリコプタ収録音の音源分離におけるシステムパラメータと分離性能について-GHDSSとBNP-MAPの比較
3. 学会等名 第35回日本ロボット学会学術講演会
4. 発表年 2017年

1. 発表者名 山田健志郎, 公文誠
2. 発表標題 Grid based Recursive Bayes Filterに基づくマルチロータヘリコプタによる音源探査における地図管理
3. 学会等名 第35回日本ロボット学会学術講演会
4. 発表年 2017年

1. 発表者名 Kazuhiro Nakadai, Makoto Kumon, Hiroshi G. Okuno, Kotaro Hoshiba, Mizuho Wakabayashi, Kai Washizaki, Takahiro Ishiki, Daniel Gabriel, Yoshiaki Bando, Takayuki Morito, Ryosuke Kojima, Osamu Sugiyama
2. 発表標題 Development of Microphone-Array-Embedded UAV for Search and Rescue Task
3. 学会等名 International Conference on Intelligent Robots and Systems (国際学会)
4. 発表年 2017年

1. 発表者名 谷口 亮輔, 干場 功太郎, 中臺 一博
2. 発表標題 Bi-directional LSTM を用いた音楽音響信号修復法の提案
3. 学会等名 第35回日本ロボット学会学術講演会
4. 発表年 2017年

1. 発表者名 高尾 麻衣子, 干場 功太郎, 中臺 一博
2. 発表標題 可聴音を用いた周波数選択に基づく距離推定法の検討
3. 学会等名 第35回日本ロボット学会学術講演会
4. 発表年 2017年

1. 発表者名 谷口 亮輔, 干場 功太郎, 中臺 一博
2. 発表標題 Quad-directional LSTMを用いた音楽音響信号修復法の提案
3. 学会等名 人工知能学会 第49回AIチャレンジ研究会
4. 発表年 2017年

1. 発表者名 高尾 麻衣子, 干場 功太郎, 中臺 一博
2. 発表標題 可聴音を用いた周波数選択に基づく距離推定法の実環境利用に向けた評価
3. 学会等名 人工知能学会 第49回AIチャレンジ研究会
4. 発表年 2017年

1. 発表者名 中臺 一博
2. 発表標題 ロボット聴覚オープンソースソフトウェアHARK の技術紹介とその展開
3. 学会等名 自動車技術会 エレクトロニクス部門 (招待講演)
4. 発表年 2017年

1. 発表者名 谷口 亮輔, 干場 功太郎, 中臺 一博
2. 発表標題 Quad-directional LSTMを用いた音楽音響信号修復とその評価
3. 学会等名 第80回情報処理学会全国大会
4. 発表年 2018年

1. 発表者名 高尾 麻衣子, 干場 功太郎, 中臺 一博
2. 発表標題 可聴音を用いた周波数自動選択に基づく距離推定法の検討
3. 学会等名 第80回情報処理学会全国大会
4. 発表年 2018年

1. 発表者名 Daniel Gabriel, Ryosuke Kojima, Kotaro Hoshiba, Kazuhiro Nakadai
2. 発表標題 Evaluation of 2D bird localization algorithm using microphone arrays
3. 学会等名 The 80th National Convention of IPSJ
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	中臺 一博 (Nakadai Kazuhiro) (70436715)	東京工業大学・工学院・特任教授 (12608)	