

ドローン自己雑音のビームフォーミング解析に基づく 地表材質推定の検討

Groud Surface Material Estimation Based on Beamforming Analysis of Drone Self-Noise

矢野 翼^{1*} Benjamin Yen^{1,2}
芦澤 剛¹ 中臺 一博¹

¹ 東京科学大学

¹ Institute of Science Tokyo

² 理化学研究所

² RIKEN

Abstract: 本稿では、ドローン自身が発するローターノイズを活用して、ドローン直下の地表材質を推定する手法を扱う。従来のデータセットは、環境の多様性や収録条件に制約があり有効なデータが限られていたため、本研究では砂利や雪を含む6種類の地表材質に対して約3時間分の音響データを新たに収録し、データセットの拡張を行った。さらに、地表方向へのビームフォーミングの適用可能性を検討し、深層学習モデルによる材質分類を行った。本研究は、ローターノイズを抑圧すべき雑音としてではなく有用な信号として積極的に活用するという、新たなドローン聴覚のアプローチを提示し、災害現場など屋外環境における堅牢なセンシング技術への貢献を目指すものである。

1 はじめに

近年、豪雨や地震などの自然災害が頻発し、土砂崩れや河川の氾濫によって、道路の寸断や孤立集落の発生といった深刻な被害が生じている。2019年の台風19号では、東日本を中心に20都県で950件を超える土砂災害が発生し[1]、2020年の熊本豪雨では、国道の複数個所で通行止めとなった。2024年の能登半島地震では、県管理の道路だけで90か所以上が通行不能となり、孤立集落の把握にも時間を要した[2]。こうした災害時には、道路や地表の状態を迅速かつ広域に把握することが復旧支援に不可欠である。特に、道路が泥に覆われているか、アスファルトが露出しているかといった地表の材質の違いは、通行可否や被害の深刻度を判断する上で重要な情報となる。しかし、従来の視覚ベースの画像認識手法では、泥濘や影、照度変化の影響により判別が困難な場合がある。そこで本稿では、音響情報を用いた地表材質推定に着目する。特に、ドローンに搭載したマイクロホンで収録する音響信号には、ローターからの直接音に加え、地表で反射した反射音も含まれており、この反射音には地表材質に応じた特徴が現れる。本稿では、この反射音を活用した地表

材質推定方法を提案するとともに、屋外環境における推定性能を評価するためのデータセットの構築も行う。

2 関連研究

災害現場における地表材質の把握は、被害状況の評価や復旧計画の策定において重要であるが、視認性やアクセス性の制約から、従来の視覚ベースの方法では限界がある。そこで、近年ではドローンにセンサを搭載し、被災状況を把握する研究が進められており、特にカメラやマイクロホンを用いた手法が提案されている[3]。カメラは、有効な観測手段だが、夜間や煙、霧など視界不良下では利用が困難という課題がある。これに対し、マイクロホンを用いた音響センシングは、視界に依存せず環境情報を取得できる利点から、「ドローン聴覚」として注目され、多くの研究が行われている[4, 5, 6, 7, 8, 9]。ただし、既存研究の多くは、被災者の音声検出や位置推定を目的としており、ローターノイズは抑圧すべき対象とされてきた。一方、本稿では、従来ノイズとみなされていたローターノイズに含まれる地表からの反射音に着目し、それを積極的に活用する点で、従来研究と一線を画す。図1に示すように、反射特性の違いから地表材質の変化を検出できれば、災害による地表面の変化（例：土砂の堆積、道路の損壊など）を把

*連絡先：東京科学大学
152-8552 東京都目黒区大岡山 2-12-1
E-mail: yano@ra.sc.eng.isct.ac.jp

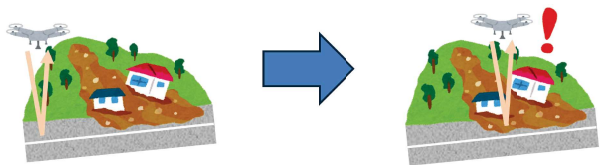
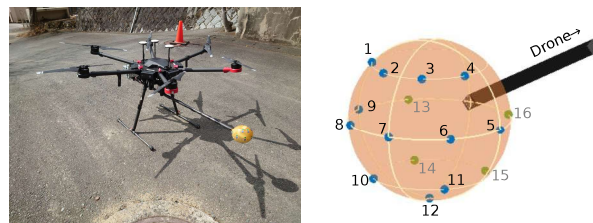


図 1: 本稿の応用例のイメージ：反射音から地表面材質を推定する。既存地図と材質が異なる部分があれば、被災により地表面に何らかの影響あったことが分かる。握できる。このようなアプローチは、われわれの知る限りでは未踏の研究領域である。

本稿では、反射音が地表面方向から到来する点に着目し、信号処理分野におけるビームフォーミング技術 [10] を適用することで反射音特徴のみを抽出し、性能向上を図る。ビームフォーミングは、複数のマイクロホンから得られる信号に時間遅延や重みづけを施すことで、特定方向に対するビームを形成し、その方向から到来する音を強調する技術であり、音源定位、音源分離、ノイズ抑圧など様々な応用が実用化されている [11]。地表面からの反射音が特定方向から到来する状況では、その方向にビームを形成することで、反射成分を選択的に抽出・強調することが可能となる。

また、音響信号を用いた材質推定の事例としては、橋梁の打音検査があげられる。従来は、熟練作業者が橋梁を打撃して音を聞き分けていたが、近年では、これを機械学習で代替する研究が進んでおり [12]、一部では、打音検査機能を有する付着走行型ドローンも報告されている [13]。ただし、これらは、主にコンクリート構造物を対象としており、多様な地表材質を対象に、反射音のみから材質を識別する研究はほとんどない。また、打音検査では、人工的に音を発生させる必要があるが、本稿では、ドローンローターから自然に発生するノイズを情報源として活用しており、従来手法とは根本的に異なるアプローチをとっている。このように、本研究は従来「抑圧すべき雑音」とされていたローターノイズを「有用な情報源」と再定義し、災害対応や野外センシングにおける音響情報の新たな活用の可能性を拓くものである。

われわれは以前、屋外環境でドローン飛行中の音響データを収録し、深層学習により地表材質を分類可能であることを示した [14]。しかし、当時のデータセットは強風の影響で有効なデータが限られ、クラス数・量ともに不十分だった。さらに、各材質につき 1 回のフライトのみの収録であったため、反射音よりも、各フライトに特有のノイズ成分を学習してしまう可能性もあった。そこで本稿では、データセットの拡張を行うとともに、データの処理が行いやすいような、外部ノイズの少ないデータ取得を狙う。また、各素材に対して複数回のフライトを行い、汎化性能の向上を図るほか、前述のビームフォーミング技術を地表材質推定手



(a) 機器取り付け後のドローン (b) アレイチャンネル配置

図 3: 収録に用いるドローンとマイクアレイ

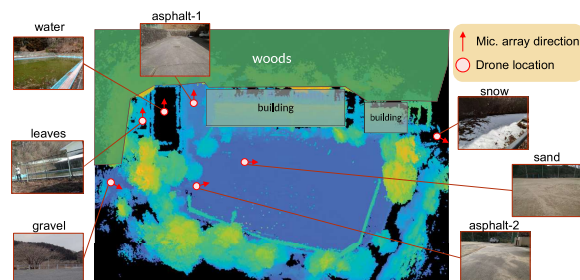


図 4: 各収録場所と地表面の様子

法に取り込むことにより、外部環境に対する頑健性の向上を狙う。

3 データセットの拡張

本稿では、地表材質推定に用いる音響データの信頼性と汎化性能を高めるため、新たな屋外データセットの収録を行った。本章では、その収録方針、手法、得られたデータの分析結果について述べる。

3.1 新規データ収録の指針

マイクロホンで録音されるローターノイズには、地表の材質だけでなく、飛行高度、各モーターの回転速度、地形など、様々な要素が影響を与える。本稿では、既存のデータセットが抱えていた課題を踏まえ、以下の方針に基づいて新たなデータを収録した。

1. **パラメータの固定化**：地表材質とドローン高度以外の要素（例：ドローンの機体設定や環境条件）を可能な限り統一する。
2. **材質の選定**：災害現場を想定し、現実的に出現しうる地表材質を選択する。
3. **高度の設定**：高度が高すぎると反射音が減衰し、推定が困難になる。風の影響も大きくなるため、人の頭上よりやや高い程度（3～5m）に限定。
4. **データ量の確保**：バッテリーの制約内で、各材質に対してできる限り長時間の収録を行う。
5. **複数回収録による汎化性向上**：1 つの地表材質につき、複数回のフライトを行う。

表 1: 新規収録データ内訳

収録場所	収録時間長 (フライト回数)
砂	79 分 (6)
アスファルト 1	19 分 (2)
アスファルト 2	20 分 (2)
水	20 分 (2)
砂利	20 分 (2)
落ち葉	20 分 (1)
雪	9 分 (1)

表 2: データセット作成条件

使用機体	DJI Matrice 600
サンプリング周波数	16 kHz

3.2 データ収録手法

収録に用いた機体について、先行研究 [14] と同様、図 3 に示すように、16ch の球形マイクロホンアレイをドローンの前方向に突き出す形で設置し、後方には、録音装置として Raspberry Pi を搭載した。機体の重心を考慮して、それぞれの位置は慎重に調整・固定した。この機体をホバリング状態にして、地表の材質を都度変えながら、取り付けたマイクロホンアレイの全てのチャンネルで録音を行った。収録対象の地表材質は、砂、アスファルト、水、砂利、落ち葉、雪の 6 種類とし、アスファルトについては周辺環境と広さが異なる 2 地点（アスファルト 1 と 2）を用意して、環境変化に対する汎化性能を評価可能にした。これらの地表材質を有する地点はいずれも大きな凹凸はなく、傾斜に関しても雪以外はほぼ水平であった。各収録地点の配置と、地表面の様子を図 4 に示す。図中の色は標高の相対差を示しており、青は低地、黄色は高地（木立や人工物のある場所）を表す。また、図中の小円はドローンの位置であり、赤い矢印の方向がマイクアレイの向きである。これらは風による影響を抑えるため、いずれも風下側を向くように設定した。収録は、各材質の上空でホバリングしながら、10 分間、もしくは 20 分間実施した。ただし、気象やドローンのバッテリー状況によって、各回の収録時間には数分程度のばらつきがある。収録時の風速は、平均して 3m/s 程度であった。最終的に約 3h 分のデータを収録することに成功した。その詳細を表 1 に、その他の条件を表 2 にまとめている。

3.3 録音データの分析

収録された音響データの品質と一貫性を評価するため、既存データと新規データの比較分析を行った。具体的には、録音中に生じる不具合の有無や、チャンネル間での音響特性のばらつきを調査し、今後の推定精度向上に向けた課題と改善点を明らかにする。

3.3.1 不具合発生の確認

既存の屋外環境データには、「音割れ」と、「録音値が 0 になる」といった不具合が含まれていた。本稿では、新規に収録したデータと既存データを比較し、こうした問題の発生頻度を調べ、表 3 にまとめた。表は、各収録場所での録音データについて、全フライトでチャンネルごとに単位時間での不具合観測回数を求め、その最大値を代表値として記載している。この結果、音割れは一部の地点で従来の 10 倍以上の頻度で発生していた一方で、録音値が 0 となる区間は大幅に減少していた。音割れの多くは、風が吹き抜けやすい場所（例：砂、アスファルト 2）で発生しており、上空では地上よりも強い風が吹いていたこと予想される。逆に木立や建物に囲まれた場所（アスファルト 1、落ち葉）では比較的発生頻度が低かったため、強い風が吹き抜けにくかったと考えられる。また、録音値 0 の区間が減少した理由としては、マイクアレイのオーバーホールにより、録音値が 0 となる現象の主要因であった振動板が内壁に貼り付く頻度が少なくなったためと考えられる。

3.3.2 チャンネル間の比較

既存のデータセットでは、各素材につき 1 回のフライトのみでデータを収録していたため、ネットワークの学習時に、複数のチャンネルのデータをまとめて入力すると、学習したい地表面からの反射音よりも、そのフライト固有の環境要因が学習されてしまう懸念があった。そこで本稿では、新たに収録したデータに対して、チャンネル間の音圧レベルを比較し、同一材質に対する録音の一貫性を検証した。図 5 は、アスファルトの上空で収録した 4 つのデータにおける、最前方の第 7 チャンネルを基準とした音圧差を示している。理想的には、第 7 チャンネルがドローンローターから最も遠いため、音圧も最も低くなると考えられるが、アスファルト 1 では機体から見て右前方の音圧レベルが低くなるなど、アスファルト 2 とは異なる傾向がみられた。これは風向きや周囲の建物など、環境の違いによる影響と考えられる。これは、すべてのチャンネルのデータをまとめて入力すると、ネットワークがフライト固有の環境要因を学習してしまうという前述の仮説を支持する結果である。したがって、今後は反射方向に着目したビームフォーミングなどの音処理技術の活用が必要であるといえよう。

4 地表材質推定手法

前節の分析結果から、チャンネルごとの音圧レベルには風向や周辺環境が大きく影響し、単純なチャンネル

表 3: 不具合発生頻度比較

区分	収録場所	音割れ	データ値 0
新規	砂	214.2	0.8
	アスファルト 1	12.3	0.0
	アスファルト 2	152.3	5.6
	水	50.9	0.0
	砂利	98.7	2.2
	落ち葉	22.7	0.0
	雪	28.0	0.0
既存	土砂	0.2	32.9
	アスファルト	0.2	15.8
	浅い水	4.1	24.0
	深い水	20.5	13.1

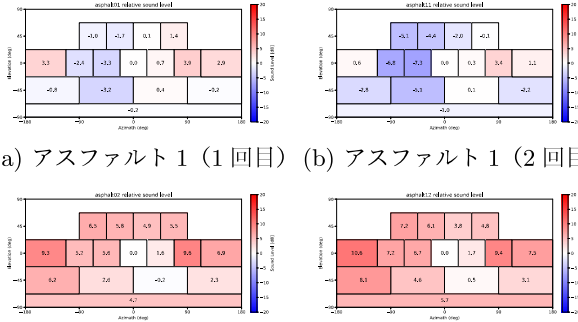


図 5: 相対音量比較

ル統合では反射音の特徴を正しく学習できない可能性が示唆された。そこで、本節では、地表からの反射音を抽出するために、録音データに対してビームフォーミングを適用する。使用する手法には、遅延和ビームフォーミング法を選択した。この手法は、各チャンネルの信号に対し、目的方向からの到来音に合わせて時間遅延を補正し、それらを加算することで、その方向からの音を強調することができる。総チャンネル数を M とし、第 m チャンネルの録音信号を x_m 、対応する時間遅延を τ_m とすると、遅延和ビームフォーミングによって抽出される目的信号 $x_{DS}(t)$ は以下で表される：

$$x_{DS}(t) = \sum_{m=1}^M x_m(t + \tau_m)$$

本稿では、方位角 180° 、仰角 -75° の方向に固定してビームを形成し、その方向に対応する時間遅延パラメータを用いた。

材質推定モデルは、先行研究 [15] を踏襲し、4 層構造の CNN (図 6) と、ResNet18 [16] を用いた。入力データ作成にあたり、ビームフォーミング後の録音信号に対してフレーム長 512、シフト長 128 でハニング窓を用いて短時間フーリエ変換 (STFT) を行い、複素スペクトログラムに変換し、直流成分を取り除く。その後、長さ 256 フレーム、シフト長 128 フレームでスペクトログラムをスライスすることで、入力データを得る。スライスした各スペクトログラムの実部と虚部

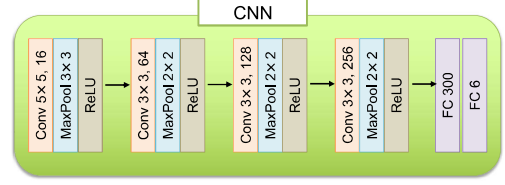


図 6: 本実験で用いる CNN アーキテクチャ

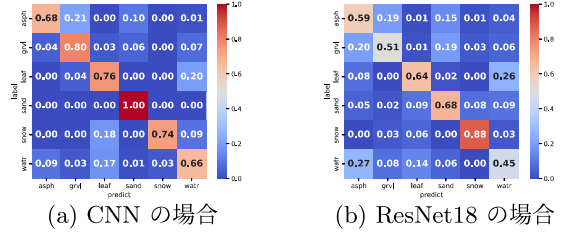


図 7: 混同行列のヒートマップ

の 2 チャンネルを入力として利用する。入力サイズは、 $2 \times 256 \times 256$ となる。

5 評価実験

構築した地表材質推定手法の有効性を確認するため、評価実験を行った。まず、各録音データに対して、遅延和ビームフォーミングを適用し、もとの録音データ中に不具合が確認されていた区間を除去した。その後、処理済み録音データを学習用、検証用、テスト用の 3 つの区間に分割し、全ての区間に対して 4 節で述べた手法により、ネットワークへの入力用の特徴量を作成した。この処理により、学習データ 4097 個、検証データ 525 個、テストデータ 536 個が得られた。学習データには、地表材質に対応する 6 次元の one-hot ベクトルを教師ラベルとして付与した。なお、表 4 に CNN と ResNet18 の学習条件を示す。学習曲線を確認し、過学習が起きていない範囲で、最も性能が良好なエポックのモデルを評価に用いた。性能指標には、以下に示す正答率 (Accuracy) を用いた。

$$\text{正答率 (\%)} = \frac{\text{推定ラベルと正解ラベルが一致したデータ数}}{\text{入力データ数}} \times 100$$

表 5 に各ネットワークにおける学習・検証・テストデータに対する正答率を示す。また、混同行列のヒートマップを図 7 に示す。CNN では、86 エポック目のモデルのテスト正答率が 70.8% となり、ResNet18 では、276 エポック目のモデルを用いて、62.3% の正答率を達成した。

実験結果の詳細を踏まえて考察を行う。CNN のテストデータに対する正答率は 70% を越えているものの、過学習が起きていない範囲では学習が収束しきれておらず、汎化性能の観点で少々疑問が残る。一方、ResNet18 はテストデータに対し 60% 程度と、CNN と比べて劣

表 4: 学習条件

損失関数	クロスエントロピー
オプティマイザ	Adam
学習率	10^{-4}
エポック数	300

表 5: 各データに対する正答率

ネットワーク	学習データ	検証データ	テストデータ
CNN	87.8%	72.2%	76.7%
ResNet18	100%	62.3%	60.2%

るものの、6 クラス分類のチャンスレートと比べて十分な高い値であることから、ある程度反射音の特徴を学習できていることが考えられる。今後はデータ区間の割り当ての変更や、既存データと混合しての検証などを通じて、汎化性能について詳しく分析を進める予定である。図 7 に示した混同行列のヒートマップを確認すると、いずれのネットワークも水に対する識別性能がほかの素材に比べて低い傾向が見られた。この原因として、プロペラによる風が水面に影響を与え、反射音の特性が不規則に変動することがあげられるが、現時点では仮説に過ぎず今後の追加分析が必要である。また、収録時に強風の影響で使用できなかったデータが存在したため、実用化に向けては、マイクロホンアレイへの風防の取り付けなど、風への対策が不可欠である。

6 おわりに

本稿では、ドローンのローターノイズに含まれる反射音を利用し、頑健な地表材質推定手法の構築を目指した。屋外環境下で、ローターノイズと地表材質 6 種類（砂、アスファルト、水、砂利、落ち葉、雪）に関するデータセットを作成し、これを対象に CNN および ResNet18 を用いた材質推定モデルを構築した。また、収録したデータに遅延和ビームフォーミングを適用したのち、地表材質推定手法の評価を行い、ResNet18 で正答率 60.2% を達成し、その有効性を示すことができた。外部ノイズや環境に対する具体的な性能の検証や、識別性能の分析、さらなるデータセットの拡張、リアルタイム処理などが、今後の課題である。

参考文献

- [1] 国土交通省 砂防部：“令和元年台風第 19 号に伴う土砂災害の概要 ver2.1”，https://www.mlit.go.jp/river/sabo//jirei/r1dosha/r1typhoon19_gaiyou191224r.pdf
- [2] 石川県：“令和 6 年（2024 年）能登半島地震に関する情報 被害等の情報について（第 12 報）”，<https://www.pref.ishikawa.lg.jp/saigai/documents/202401041500higaihou.pdf>
- [3] 松井他：“ドローン搭載カメラからの RGBD 画像に対するセグメンテーション等による土砂災害箇所と規模の推定法の検討”，研究報告コンピュータビジョンとイメージメディア (CVIM), no. 16, pp. 1-6, 2023.
- [4] K. Okutani et al.：“Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter”, IEEE/RSJ IROS, pp. 3288-3293, 2012
- [5] K. Hoshiba et al.：“Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments”, Sensors, 17(11), 2017
- [6] T. Yamada et al.：“Placement Planning for Sound Source Tracking in Active Drone Audition”, Drones, 7(7), 2023
- [7] M. Kumon et al.：“Alternating Drive-and-Glide Flight Navigation of a Kiteplane for Sound Source Position Estimation”, IEEE/RSJ IROS, pp. 2114-2120, 2021
- [8] B. Yen et al.：“A Performance Assessment on Rotor Noise-Informed Active Multidrone Sound Source Tracking Methods”, Drones, 8(6), 2024
- [9] L. Wang et al.：“Drone audition for bioacoustic monitoring”, Methods in Ecology and Evolution, 14(12), pp. 3068-3082, 2023
- [10] Van Veen, B. D. and Buckley, K. M.：“Beam-forming: a versatile approach to spatial filtering”, IEEE ASSP Magazine, 5(2), pp. 4-24, 1988
- [11] P. Chiarioti et al.：“Acoustic beamforming for noise source localization – Reviews, methodology and applications”, Mechanical Systems and Signal Processing, vol. 120, pp. 422-448, 2019
- [12] 江本他：“AI 手法による打音検査の浮き判定の検討”，AI・データサイエンス論文集, 1(J1), pp. 514-521, 2020
- [13] 上半：“ドローンを用いたコンクリート橋部材の詳細検査手法”，コンクリート工学, 57(9), pp. 699-704, 2019
- [14] 矢野他：“屋外環境下でのドローンのローターノイズによる地表材質推定に向けた手法の検討およびマイクロホンアレイ用風防の設計”，SIG-Challenge, 2024

- [15] 矢野他：“ドローンのローターノイズによる地表材質推定手法の検討”, SICE SI, 2023
- [16] K. He et al.：“Deep Residual Learning for Image Recognition”, CVPR, pp. 770-778, 2016