

敵対的映像攻撃が自律飛行ドローンの位置推定・制御に及ぼす影響評価

海老根 佑雅^{1,a)} 野本 一輝^{1,2} 田中 優奈^{1,2} 鶴岡 豪¹ 森 達哉^{1,3,4}

概要：Visual SLAM (vSLAM) は、自動運転やロボットのナビゲーションなど、様々なシーンにおいて重要な役割を果たしており、自律飛行ドローンにおいてもその重要性は増してきている。一方で、従来の研究では信頼性と精度の高い vSLAM システムの構築に焦点が当てられており、既存の vSLAM システムの脆弱性に関する研究はほとんど行われていない。本研究では、vSLAM 技術の一種である ORB SLAM3 を用いた自律飛行ドローンに対する新たな攻撃手法「敵対的映像攻撃」を提案する。本手法は、スクリーンに投影した敵対的映像を用いて、自己位置推定を本来の結果から誤った方向へ誘導し、ドローンの進路を意図的に操作するものである。本稿では、敵対的映像が自己位置推定に与える影響をシミュレータ上のカメラ及び商用のカメラで評価し、敵対的映像攻撃が ORB SLAM3 の自己位置推定に最大で約 252 m の誤認識を引き起こすことを明らかにした。また、敵対的映像がドローンの自律飛行機構に与える影響をシミュレータ及び商用ドローンを用いて End-to-End で評価し、シミュレータ上、実世界上どちらにおいても、敵対的映像攻撃により、意図的にドローンを誘導し、墜落させることができることを示した。

Evaluation of the Impact of Adversarial Video Attacks on Localization and Control of Autonomous Drones

YUGA EBINE^{1,a)} KAZUKI NOMOTO^{1,2} YUNA TANAKA^{1,2} GO TSURUOKA¹ TATSUYA MORI^{1,3,4}

Abstract: Visual SLAM (vSLAM) is a critical technology in various fields, including autonomous driving and robot navigation, and its importance for autonomous drones continues to grow. However, while conventional research has primarily focused on developing reliable and accurate vSLAM systems, the vulnerabilities of these existing systems have received little attention. In this study, we propose a novel attack method, the "Adversarial Video Attack," targeting autonomous drones that utilize ORB SLAM3. This method employs an adversarial video projected onto a screen to mislead the drone's localization, thereby intentionally manipulating its trajectory away from the intended path. We evaluated the impact of this adversarial video on localization using both simulated and commercial cameras. Our results demonstrate that the Adversarial Video Attack can induce a maximum localization error of approximately 252 meters in ORB SLAM3. Furthermore, we conducted an End-to-End evaluation of the attack's effect on the drone's autonomous navigation system using both a simulator and a commercial drone. We successfully demonstrated that in both simulated and real-world environments, the Adversarial Attack can be used to intentionally steer the drone off its course and ultimately cause it to crash.

1. はじめに

Visual Simultaneous Localization and Mapping (vSLAM) は、カメラへの入力映像から正確な自己位置推定と地図作成を可能にする技術であり、自律飛行ドローンの飛行において重要な役割を果たす。vSLAM は、インフラ点検、捜索救助、自律探査といった用途で活用され始めている [1]。

しかし、vSLAM は視覚データに大きく依存するため、カメラセンサに対する敵対的な挿動に対して脆弱であり、その結果生じる自己位置推定の誤差は、安全性やセキュリティ上の深刻なリスクをもたらす可能性が指摘されている [2], [3]。

¹ 早稲田大学/Waseda University

² デロイト トーマツ サイバーコンソーシアム/Deloitte Tohmatsu Cyber LLC

³ 国立研究開発法人情報通信研究機構/NICT

⁴ 理化学研究所 革新知能統合研究（AIP）センター/RIKEN AIP

a) yuga@nsi.cs.waseda.ac.jp

近年の研究では、vSLAM システムの脆弱性を探るため、物理的なオブジェクトの設置や光の投影といった多様な攻撃手法が実証されている。物理的なパッチを環境内に設置する手法はその代表例である。例えば、Chen ら [2] は、視覚的特徴量を微細に変化させ、SLAM ベースの自己位置推定を著しく妨害する敵対的パッチ攻撃を提案した。Nemcovsky ら [4] は、多数の類似した特徴点を意図的に生成する模様を持つパッチを壁面などに配置し、ビジュアルオドメトリモデルの特徴点の追跡を混乱させることで誤差を大幅に増大させ、自律飛行ドローンの飛行に重大な脅威をもたらすことを示した。また、環境に光を投影する手法も存在する。Wang ら [5] は、カメラのみが検知可能な赤外線光で偽の特徴点を物体に投影し、自動運転車で利用される SLAM システムに誤差を誘発する攻撃を提案した。さらに、vSLAM システムの特定コンポーネントを標的とする研究や、ドローン自体を対象とする研究も存在する。例えば Doe ら [6] は、視覚的特徴の微細な曖昧さを悪用してループクロージャを妨害する、知覚的エイリアシングに基づく攻撃を提案しており、他にもドローンの慣性センサ等を標的とする攻撃が複数報告されている [7], [8], [9]。しかしながらこれらの研究は SLAM システムの特定のコンポーネントに対する攻撃評価に留まっており、vSLAM をもとに自律的に飛行するドローンの自己位置推定を誘導することで、その制御に誤作動を生じさせるものではない。

本稿では、敵対的に生成した特殊な白黒映像をドローンに認識させることで、自己位置推定を任意の方向に誘導し、ドローン墜落のリスクを大幅に増大させる攻撃である敵対的映像攻撃を提案する。敵対的映像攻撃は図 1 に示すように、vSLAM ベースのドローンを誤誘導ものである。具体的には、敵対的映像として、vSLAM の自己位置推定を意図的に騙す白黒模様の映像を作成し、前方方向に飛行する自律飛行ドローンに敵対的映像を認識させることで、その映像を入力とする vSLAM は偽の自己位置推定結果 (Phantom Path) を生成し、その結果ドローンの挙動に乱れを生じさせる。静的な敵対的パッチに依存する従来の手法 [4] とは異なり、本攻撃は敵対的に生成された映像投影を用いて vSLAM の自己位置推定を任意の方向に誘導し、ドローンの誤作動を引き起こすものである。

敵対的映像攻撃の影響を評価するため、本稿では、攻撃の実現可能性を評価する「カメラ単体評価」と、その攻撃が自律飛行ドローンの制御に与える影響を評価する「シミュレータ及び商用ドローンを用いた End-to-End (E2E) 評価」を行う。両実験には、vSLAM システムとして広く採用されているオープンソースの ORB SLAM3 [10] を採用する。実験の結果、カメラ単体評価では最大 252 m の自己位置推定誤差が示され、シミュレータを用いたシミュレーション上で E2E 評価では約 70 m の降下制御が墜落につながることを示した。商用ドローンを用いた実世界での E2E 評

価では、敵対的映像の投影により、ドローンを任意の方向に誘導できることを確認した。

これらの評価結果は、vSLAM ベースの自律飛行ドローンにおける重大な脆弱性を明らかにしている。この脅威を軽減するため、我々は LiDAR や IMU データを統合するセンサフュージョン技術や、動的な物体のフィルタリングを含む、潜在的な防御策を提案する。具体的には、センサフュージョンによる偽の移動情報の検知や、カメラへの入力から動的物体をフィルタリングすることで敵対的映像を入力映像から除外することを提案する。

本研究の貢献を以下に示す。

- 図 1 に示すように、映像の投影を利用して vSLAM ベースの自己位置推定を操作する敵対的手法である敵対的映像攻撃を提案する。
- 本攻撃が ORB SLAM3 において最大 252 m の深刻な自己位置推定誤差を誘発することを示す。
- vSLAM システムの自己位置推定結果に基づいて飛行する自律飛行ドローンへの攻撃に対する初の E2E 評価を行い、シミュレータ上、実世界上の両方において、任意の方向に自律飛行ドローンの制御を誘導できることを明らかにする。
- vSLAM ベースの自律飛行システムのセキュリティを強化するため、センサフュージョンと動的物体の除去を用いた防御手法を提案する。

2. 背景と関連研究

2.1 自律飛行ドローンと vSLAM

自律飛行ドローンは、外部からの指示なしに、搭載されたセンサを用いて自己位置推定や経路計画を行いながら動作し、災害救助、農業、物流、インフラ点検など、幅広い分野で利用されている [11]。しかし、屋内環境や GNSS 信号が利用できない複雑な都市部では、自己位置推定が困難になる。

この課題を克服するため、軽量でカメラベースの自己位置推定手法として vSLAM が用いられるようになった [1]。vSLAM は、特徴点の抽出とマッチング、姿勢推定、地図構築を行うことで、広範囲の環境マッピングと自己位置推定を同時に実現する。この機能は特徴点抽出、特徴点マッチング、姿勢推定、地図構築から構成される。代表的なアルゴリズムとして ORB SLAM3 [10], LSD-SLAM [12], PTAM [13], MonoSLAM [14] などがあり、中でも ORB SLAM3 はその精度の高さと効率性で広く採用されている。

2.2 ORB SLAM3 の概要

ORB SLAM3 は、ORB 特徴量を用いてリアルタイムの自己位置推定とマッピングを可能にする vSLAM アルゴリズムであり、GNSS が利用できない環境で動作する自律飛

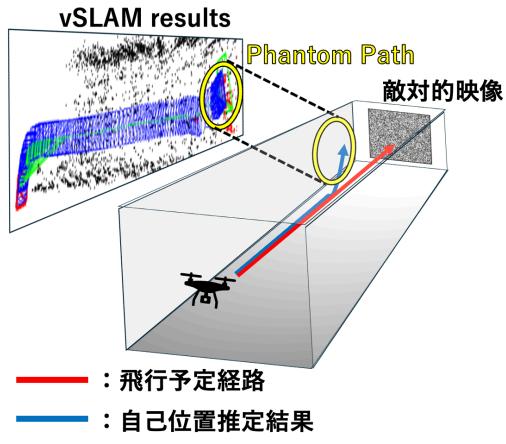


図 1: 攻撃の概要図

行ドローンに適している。このアルゴリズムは主にトラッキング、ローカルマッピング、ループクロージャの3つの重要なステップで構成される。トラッキングでは、FASTコーナー検出と BRIEF 記述子を用いて ORB 特徴量を抽出し、フレームごとにカメラの動きを推定する。ローカルマッピングでは、キーフレームを三角測量して 3D マップを構築し、バンドル調整によって構造を精密化する。ループクロージャでは、特徴点マッチングによって過去に訪れた場所を検出し、PnP とバンドル調整を用いてグローバルマップを最適化し、ドリフト（誤差の蓄積）を補正する。

ORB SLAM3 では、入力映像から抽出された特徴点の対応付けに失敗すると、システムは自己位置推定と環境地図の構築能力を失う。この現象はトラッキングロスと呼ばれる。ORB-SLAM3 は単眼カメラ、ステレオカメラ、RGB-D カメラに対応しており、これを用いることでドローンは複雑な環境でも高精度な航行が可能となる。また、安定した自己位置推定とドリフト補正能力は、インフラ点検、捜索救助、自律探査といった用途において極めて有用である。

3. 攻撃の概要と脅威モデル

3.1 攻撃の概要

本研究では、vSLAM の自己位置推定に誤認識を引き起こす敵対的映像を用いることで、攻撃対象のドローンを任意の方向に誘導する敵対的映像攻撃を提案する。攻撃の概要を図 1 に示す。この攻撃では ORB SLAM3 の特徴点抽出アルゴリズムとそれに付随する自己位置推定の特性を利用する。攻撃者は、ターゲットとなるドローンが自律飛行を行う建物内にスクリーンを設置し、特徴点を集めやすい白黒の幾何学模様が流れる映像を投影する。ターゲットとなるドローンに搭載された vSLAM アルゴリズムは、敵対的映像から本来は存在しない特徴点を抽出することで、誤った自己位置推定を行う。その結果、ターゲットとなるドローンは、通常とは異なる経路を飛行する。

3.2 脅威モデル

攻撃の目的: 提案する攻撃の目的は、自律飛行ドローンの自己位置推定に誤認識を引き起こし、ドローンの制御を乱すことで、任意の方向への誘導あるいは墜落を誘発することである。攻撃者は例えば巡回を行う自律飛行警備ドローンの制御を意図的に乱すことで、警備の眼を掻い潜って行動を行うことができる。

攻撃者の仮定: 攻撃者は標的の自律飛行ドローンが飛行する予定のルートを把握し、自律飛行ドローンが ORB SLAM3 等の特徴点ベースの vSLAM をもとに自律飛行することを知っている必要がある。攻撃者は映像を投影できるプロジェクトを所有し、投影する映像を作成することができる能力があると仮定する。設置場所には映像を投影できるスペースが確保されている必要がある。攻撃者は遠隔から攻撃を実行することが可能である。

3.3 攻撃の手順

ステップ 1: 敵対的映像の作成: このステップでは、攻撃に使用する敵対的映像の作成を行う。攻撃者は敵対的映像のみで自律飛行ドローンの自己位置推定に誤認識を引き起こすために、輝度差の生じやすい敵対的映像を作成する必要がある。また、ORB SLAM3 のトラッキングロスを予防するため、映像は特定の模様の繰り返しにより構成されていない必要がある。なお、敵対的映像の生成メカニズムの詳細は、3.4 節に示す。

ステップ 2: 実世界での映像源の設置: このステップでは、攻撃者はその映像を実世界で投影するためにプロジェクトを任意の位置に設置する。壁面等に映像を直接投影できないような状況ではスクリーンも設置する必要がある。この時、標的ドローンが通るルートを攻撃者自身が把握しておく必要がある。設置場所は必ずしもオープンスペースである必要はなく、標的ドローンが飛行することがわかっている場所であれば、任意の場所に設置して問題ない。

ステップ 3: 敵対的映像の投影: このステップでは、攻撃者はプロジェクト及びスクリーン設置後に映像を投影するタイミングを決定する。自律飛行ドローンが映像設置場所の付近を通るタイミングを見計らって映像を投影する。一方で、敵対的映像のループ再生を行うことで、厳密にドローンの飛行タイミングを見計らう必要はなくなる。

3.4 敵対的映像

FAST アルゴリズム: ORB SLAM3 では入力された映像の各カメラフレームに対して、FAST アルゴリズムによりコーナーを検出し、それを後に ORB 特徴点として抽出する。FAST アルゴリズムでは各カメラフレーム内の各ピクセルを対象として、その周囲 16 ピクセルとの輝度差の差分を計算し、閾値を超えて明るいあるいは暗いピクセルが 9 連続以上あれば、対象のピクセルをコーナー候補として残す。

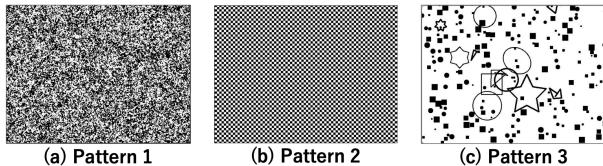


図 2: 敵対的ビデオを構成する 3 種類のパターン

表 1: 各映像における FAST アルゴリズムを用いた 900 フレームあたりのコーナー検出数

Pattern 1	Pattern 2	Pattern 3
10230973	1831014	1421099

コーナー候補のうち、近傍のコーナーどうしを比較して局所最大なコーナーのみを残し、それを特徴点とする。

敵対的映像の生成: 敵対的映像の生成において、自己位置推定への干渉とトラッキングロスの防止の 2 つの観点が重要である。まず、自己位置推定への干渉に関して、敵対的映像から効率的に特徴点が抽出される必要がある。すなわち、敵対的映像から FAST アルゴリズムに基づき、効率的にコーナーが検出される必要がある。すなわち、敵対的映像はより多くピクセル間の輝度差を生むため、白黒を基調とした映像である必要がある。続いて、トラッキングロスの防止に関して、ORB SLAM3 はチェック模様のような同じ模様の繰り返しを認識すると、各特徴点の特徴の対応づけに失敗し、自己位置推定及び環境地図の作成を停止してしまう。そのため、敵対的映像は一定程度ランダムな模様により構成されている必要がある。以上の観点に基づき、本研究では敵対的映像の作成にフラクタルノイズを利用する。具体的には、異なる解像度で生成されたノイズを加算することで異なるスケールの構造を持つフラクタルノイズを生成し、Gaussian ブラーと閾値処理により、図 2 に示す模様で構成される Pattern 1 を敵対的映像として作成する。また、図 2 に示される模様で構成される Pattern 2 や Pattern 3 も用意し、攻撃性能の比較を行う。

コーナー検出数の比較: 用意した 3 種類の映像に対して FAST アルゴリズムを用いて映像のコーナー検出をそれぞれ 5 回ずつ行い、その平均値を比較した結果を表 1 に示す。この結果から図 2 に示す Pattern 1 の模様で構成される映像が最もコーナーが検出される映像であることを確認した。

4. カメラ単体を用いた基礎評価

本章は、実際に作成した 3 つの敵対的映像を用いて、その映像が ORB SLAM3 の自己位置推定に与える影響を評価する。提案攻撃が与える影響を網羅的に評価するため、シミュレーション評価と実世界評価を行う。シミュレーション評価では、仮想空間を用いることで、攻撃手法の基本的な有効性や攻撃後の挙動特性を定量的に評価する。シミュ

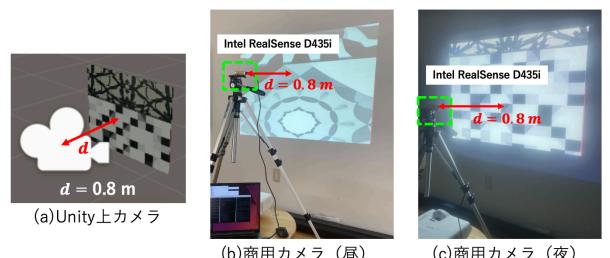


図 3: カメラ単体評価における実験環境

レータとして Unity を利用し、Unity 上に設置したカメラの映像を入力として、ORB SLAM3 が output する自己位置推定の結果を評価する。実世界評価では、市販のカメラおよびプロジェクトを用いて現実的な環境条件下で攻撃の適用可能性を検証する。実験では、一般的に販売されるカメラやプロジェクトを用いて攻撃を実施し、ORB SLAM3 が output する自己位置推定の結果を評価する。

4.1 シミュレータを利用した評価

実験セットアップ: 本実験では、シミュレータとして Unity を用い、Unity 内の仮想空間において攻撃を再現することで、提案攻撃の実現性を評価する。実験環境を図 3 の (a) に示す。仮想空間上に敵対的映像を投影するスクリーンが設置され、スクリーンに敵対的映像が投影される。敵対的映像として、3.4 節で生成した映像を使用し、スクリーンには自己位置推定が上方向に誘導される向き、すなわち映像が下方向に流れしていく向きに映像を投影する。また、映像が投影されるスクリーンの前方 $d = 0.8, 0.9, \dots, 1.3$ [m] に攻撃対象となるカメラが固定される。ここで、カメラの内部パラメータは 4.2 節の実世界評価で用いるカメラ Intel RealSense Depth Camera D435i に基づいて設定される。カメラで取得された映像は、ROS2 内で動作する ORB SLAM3 に入力され、vSLAM による自己位置推定やマップ作成が行われる。

実験手順: 実験では、シミュレータ上のスクリーンに敵対的映像が投影され、シミュレーション内に配置されたカメラが敵対的映像を含む映像を 30 秒間撮影する。撮影された映像は、ROS2 空間内で動作する ORB SLAM3 に入力される。このとき、ORB SLAM3 が output する自己位置推定結果を計測する。30 秒経過後に、ORB SLAM3 による位置推定結果を取得し、終了後にその移動距離を算出する。

実験結果: 非攻撃時の ORB SLAM3 の出力結果を図 5 に示す。続いて、攻撃時の各距離ごとの ORB SLAM3 の出力結果と、自己位置推定の移動距離をそれぞれ図 ??、表 2 に示す。この結果はカメラとスクリーン間の距離 d の違いに起因する、スクリーンがカメラ画角に占める割合によって、提案攻撃が ORB SLAM3 の自己位置推定に与える影響が異なることを意味している。Pattern 1 を使用し、カメラとスクリーン間の距離 d が 0.8 m の時、移動距離が約 125 m と

表 2: シミュレータ環境におけるスクリーンとカメラ間の距離、および各映像の自己位置推定のずれ(単位: m)

スクリーンとカメラ間の距離 d	Pattern 1	Pattern 2	Pattern 3
0.8	124.7	123.6	38.7 *
0.9	114.9	115.2	115.2
1.0	96.8	98.7	104.8
1.1	81.0	92.1	88.0
1.2	12.7 *	84.1	92.6
1.3	18.3 *	41.3	12.5 *

* ラッキングロスが生じた

なり、攻撃の有効性が最も高くなったことを示している。

この結果は、敵対的映像から効率的に特徴点が抽出され、かつ、ラッキングロスすることなく抽出された特徴点が自己位置推定に用いられたことによるものである。この結果は敵対的映像を認識させることで自己位置推定を大幅に任意の方向に移動させることができることを示している。

4.2 実世界評価

本実験では、市販のカメラおよびプロジェクタを用いて、現実的な環境条件下で攻撃の適用可能性を検証する。

実験セットアップ: 実世界評価における実験環境を図 3 の (b), (c) に示す。商用カメラとして Intel RealSense D435i, プロジェクタとして EPSON EB-L200SW を使用する。各セットアップは、4.1 と同様に、敵対的映像が投影されたスクリーンの前方 $d = 0.8, 0.9, \dots, 1.3$ [m] にカメラを固定して撮影する。また、スクリーンには自己位置推定が上方に向かって誘導される向き、すなわち映像が下方に向かって流れいく向きに映像を投影する。取得されたカメラ映像は ORB SLAM3 に入力され、自己位置推定結果が記録される。なお、実験では周囲の環境が攻撃に与える影響を評価するため、明るい環境（昼環境）と暗い環境（夜環境）で実験を行う。

実験手順: 本実験では実世界上における攻撃時の ORB SLAM3 の出力を検証する。攻撃実験では、実世界上の白い壁に横 0.9 m, 縦 1.2 m の敵対的映像を投影する。4.1 節と同様に、カメラを敵対的映像から $d = 0.8, 0.9, \dots, 1.3$ [m] の距離に配置し、30 秒間撮影し ORB SLAM3 に入力する。いずれの実験においても、ORB SLAM3 が出力する自己位置推定結果を取得し、最初と最後の座標情報から、自己位置推定の移動距離を算出する。

実験結果: 攻撃時の昼間および夜間における各距離ごとの ORB SLAM3 出力結果を、それぞれ図 ?? に示す。また、各動画・条件ごとの自己位置推定の移動距離を表 3 に示す。これらの結果はシミュレータ上のカメラを利用した評価

表 3: 実世界評価におけるスクリーンとカメラ間の距離、及び各映像の自己位置推定のずれ(単位: m)

スクリーンと カメラ間の 距離 d	Pattern 1		Pattern 2		Pattern 3	
	昼	夜	昼	夜	昼	夜
0.8	231.1	227.5	-	-	243.6	251.7
0.9	204.1	221.7	-	-	216.3	219.2
1.0	199.8	195.3	-	-	207.8	212.8
1.1	176.9	175.4	-	-	171.5	186.9
1.2	162.2	155.4	-	-	171.8	157.4
1.3	152.3	145.3	-	-	156.3	123.7

- ORB SLAM3 は起動しなかった

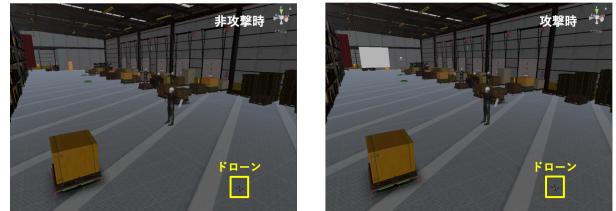


図 4: シミュレータを利用した E2E 評価における実験環境

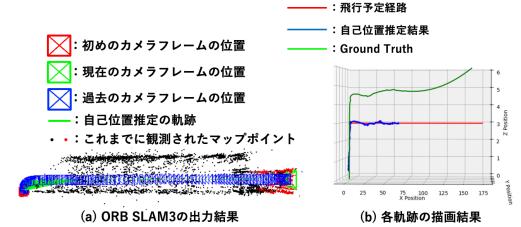


図 5: シミュレータを利用した E2E 評価における非攻撃時の側面から見た ORB SLAM3 及び各軌跡の出力結果

と同様に、スクリーンとカメラ間の距離によって、提案攻撃が ORB SLAM3 の自己位置推定に与える影響が異なることを意味している。夜にスクリーンとカメラ間の距離が 0.8 m の時に Pattern 3 を使用した際、移動距離が約 252 m と最大になり、攻撃の有効性が最も高いことが示された。

5. シミュレータを利用した E2E 評価

本章では、カメラ単体を利用した実験にて有用性を確認した敵対的映像による攻撃が、自律飛行ドローンの挙動に及ぼす影響をシミュレータを用いて E2E で評価する。

5.1 実験セットアップ

本実験では FAST アルゴリズムにより最もコーナーが検出された Pattern 1 を攻撃用の敵対的映像として用いる。また、攻撃時及び非攻撃時の実験環境を図 4 に示す。シミュレータには引き続き Unity を使用し、倉庫のアセットを使用して屋内飛行シナリオについての検証を行う。ドローン

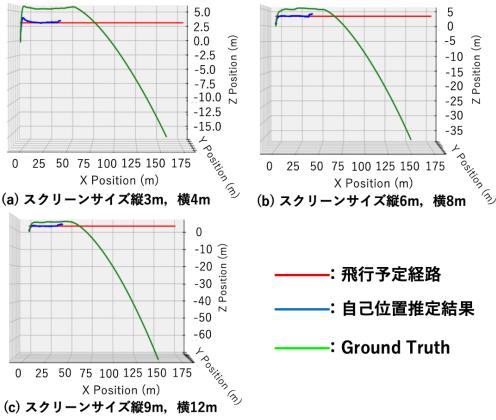


図 6: シミュレータを利用した E2E 評価における攻撃時の側面から見た各軌跡の出力結果

はスクリーン正面の地上 0 m に設置し、ドローンとスクリーンの間の距離を 60 m に固定する。スクリーンには自己位置推定が上方向に誘導される向き、すなわち映像が下方向に流れていく向きに映像を投影する。シミュレーションが開始すると離陸とともに Pure Pursuit アルゴリズムに基づき PD 制御が行われる。なお、Look Ahead Distance は 0.5 m に設定されている。Pure Pursuit アルゴリズム内ではドローンの飛行予定経路と自己位置推定結果を比較し、自己位置推定結果が飛行予定経路に沿うようにドローンを制御する。飛行予定経路は 175 個の Waypoint で構成され、それらの Waypoint は地上から高さ 3 m に 1 m 間隔でドローンの前方方向にそれぞれ設定されている。今回はドローンの動作モデルとして近似的なモデルを採用し、x 軸、y 軸、z 軸方向にそれぞれ独立の加速度を付与する形式での制御を行う。また、Pure Pursuit アルゴリズムではドローンの上下左右方向の位置調整のみを考慮し、前進運動をデフォルトで仮定する。デフォルトで規定する前進運動では、前進方向に 1.0 m s^{-2} の加速度を常に加え続けるが、速度制限を設け、最大 5.0 m s^{-1} の速さで前進し続ける。まず、スクリーンを設置しない非攻撃時の自律飛行ドローンの挙動を確認する。次に、ドローンの前方に敵対的映像を映し出すスクリーンを設置し、スクリーンのサイズごとにドローンの挙動を評価する。スクリーンのサイズは 3 種類用意し、それぞれ (I) 縦 9 m、横 12 m、(II) 縦 6 m、横 8 m、(III) 縦 3 m、横 4 m である。ドローンは時間経過とともに前進していく、一定時間経過後にカメラ画角内に敵対的映像を捉えることで、自己位置推定に誤認識が生じる。

5.2 実験手順

実験は、シミュレータ内に敵対的映像が投影されない非攻撃実験と、敵対的映像が投影される攻撃実験からなる。いずれの実験においてもドローンを初期位置に設置し、ORB SLAM3 の自己位置推定結果と事前に与えられた Waypoint をもとに、Pure pursuit アルゴリズムによりドロー

ンの制御が行われる。実験では ORB SLAM3 の自己位置推定結果の座標情報をトピックとして収集し、計測する。シミュレーションは 60 秒間行い、その間の自己位置推定結果及びシミュレータ上のドローンの Ground Truth の軌跡を取得し、飛行予定経路の軌跡と比較する。

5.3 実験結果

非攻撃時の各軌跡の描画結果と ORB SLAM3 の出力結果を図 5 に示す。非攻撃時の結果はから、Ground Truth と自己位置推定値との間に、高さ方向で約 1~2 m のずれが生じることが示された。これは ORB SLAM3 は屋内環境において自己位置推定に 1 m 程の誤差を生じてしまう、かつ、シミュレーション内でドローンの離陸と ORB SLAM3 の起動にタイムラグがあり、わずかにオフセットが生じてしまうためであると考えられる。また、攻撃時の各軌跡の描画結果を図 6 に示す。この図は自己位置推定結果が途切れる寸前に上方向に軌跡が伸びており、また、Ground Truth が下方向に制御されていることを示している。これらの結果は、敵対的映像攻撃により自律飛行ドローンの自己位置推定が上方向に誘導され、付随して実際のドローンの挙動が下方向に制御され、ドローンが墜落する可能性があることを示している。特に、スクリーンのサイズが縦 6 m、横 8 m の時、Ground Truth の軌跡から、最終的に初期位置から 70 m 程下方向に移動していることが確認できる。また、実験結果から ORB SLAM3 が途中でトラッキングロスしていることが読み取れる。このトラッキングロスは、スクリーン通過後に認識する映像がこれまで構築してきたマップと一致しないことにより生じるものである。

6. 商用ドローンを利用した実世界 E2E 評価

本章では、カメラ単体評価にて有用性を確認した敵対的映像による攻撃が、実世界上にて自律飛行ドローンの挙動に及ぼす影響を E2E で評価する。ドローンとして DJI 製の Tello [15] を使用する。本実験では、ドローンの制御には PID 制御を用いる。この制御では、まず飛行予定経路上に設置された各 Waypoint と各時刻での自己位置推定結果を比較する。次に、その距離差分をもとに制御情報を算出し、ROS 2 を介して Tello へ送信する。実際に使用するドローンを図 7 の (a) に示す。LiDAR で効率的に追跡するため、ドローンの上部と後部には再帰反射パッチを貼り付けている。

6.1 妥当性評価

実験セットアップ: 本実験では 5.1 節と同様に、Pattern 1 を攻撃用の敵対的映像として用いる。また、妥当性評価における実験環境を図 7 の (b) に示す。本実験ではドローンの設置場所後方に LiDAR センサを設置し、ドローンの Ground Truth の取得を行う。LiDAR センサとして、Ouster

OS1-64 [16] を使用する。

実験手順: まず、ドローンを初期位置に設置する。機体は、ORB SLAM3 による自己位置推定結果と、事前に与えられた 0.5 m 間隔の 4 個の Waypoint との差分に基づき、PID 制御によって自律的に飛行する。実験では ORB SLAM3 の自己位置推定結果の座標情報をトピックとして収集し、計測する。ドローンには前方方向へ 1 m 移動し、その後 0.5 m 左方向へ移動する飛行予定経路が与えられる。その際のドローンの飛行する位置を LiDAR センサで計測し、ドローンをとらえた点群の移動情報から、ドローンが飛行した軌跡の Ground Truth を取得し、事前にトピックとして存在する飛行予定経路や、ORB SLAM3 の自己位置推定結果との比較を行う。

実験結果: ORB SLAM3 の自己位置推定結果をもとに自律飛行を行うと、飛行予定経路と実際に飛行した経路との間に最大で 1 m 程度の誤差が生じた。妥当性評価における真上から見た各軌跡を図 8 に示す。図を参照すると、特に奥行き側へ進む際に 1 m 程度の誤差が生じているが、これは奥行き方向に対して実際の移動よりも ORB SLAM3 の自己位置が短く推定されたことに起因する。

6.2 攻撃評価

実験セットアップ: 攻撃評価における実験環境を図 7 の (c) に示す。攻撃時には、スクリーンとドローン間の距離が 1 m となるようにドローンの正面にスクリーンを設置する。その後、縦 1.4 m、横 1.1 m の敵対的映像を投影した状態で自律飛行を実施する。

実験手順: 妥当性評価と同様にまず、ドローンを初期位置に設置する。機体は、ORB SLAM3 による自己位置推定結果と、事前に与えられた 0.5 m 間隔の 4 個の Waypoint との差分に基づき、PID 制御によって自律的に飛行する。実験では ORB SLAM3 の自己位置推定結果の座標情報をトピックとして収集し、計測する。また、ドローンが敵対的映像を認識した際の自己位置推定結果と Ground Truth それぞれについて、飛行予定経路からの誤差を計測し、比較検討する。

実験結果: 実験結果から、敵対的映像の投影により、実世界において自律飛行ドローンを任意の方向に誘導することが可能であることが示された。攻撃評価における各軌跡の取得結果を図 9 に示す。この結果から、ドローンは離陸後、敵対的映像を認識することで自己位置推定が右側に誘導され、その誤差を軽減するために、実際には左方向への制御がなされることが確認できる。

7. 議論

7.1 制約事項

vSLAM システムによる差異: 本研究では、代表的な vSLAM システムの 1 つである ORB SLAM3 を評価対象として選定

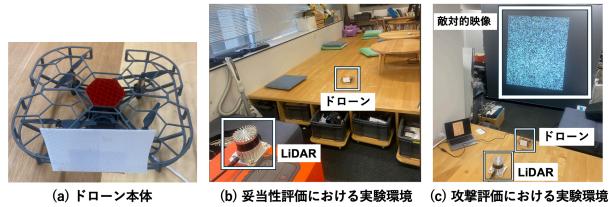


図 7: 実験で使用するドローン本体と妥当性評価及び攻撃評価における実験環境

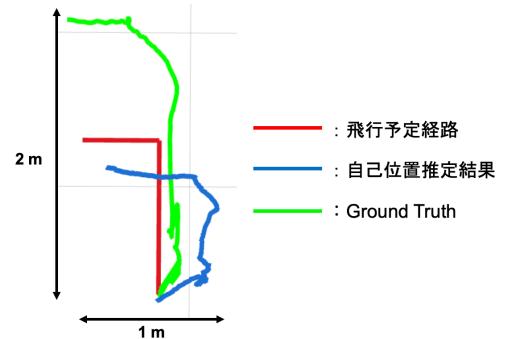


図 8: 妥当性評価における、真上から見た各種軌跡の描画結果

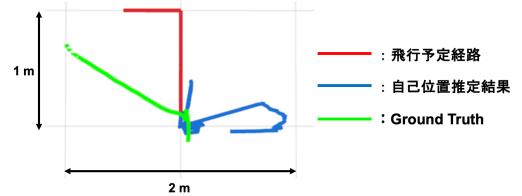


図 9: 攻撃評価における、真上から見た各種軌跡の描画結果

し、攻撃の有効性を確かめた。しかし、使用する vSLAM システムによって、攻撃の有効性が異なることが想定される。ORB SLAM3 とは異なる特徴点抽出アルゴリズムを採用する vSLAM システムにおける攻撃の有効性の評価は今後の課題である。

制御アルゴリズムの制約: 本研究では、シミュレータ上のドローンの制御アルゴリズムとして Pure Pursuit、実世界上の商用ドローンの制御アルゴリズムとして PID 制御を使用した。しかし、Pure Pursuit アルゴリズムや PID 制御は、すべてのドローンで採用される仕組みではなく、異なる制御アルゴリズムを採用する自律飛行ドローンの開発や販売が想定される。自律飛行ドローンに用いられる他の制御アルゴリズムに対する攻撃の影響評価は、今後の課題である。

シミュレータ上のドローン動作モデル: 本研究のシミュレーション内で使用した自律飛行ドローンの動作モデルは近似的なモデルであり、必ずしも実世界上の自律飛行ドローンの挙動を反映していない。動作モデルが実世界上の機体に即したシミュレータ環境内の自律飛行ドローンに対する敵対的映像攻撃の有効性評価は今後の重要な課題である。今後は実世界上のドローンの動作モデルを利用して、

敵対的映像攻撃の有効性評価を行う予定である。

攻撃のステルス性: 敵対的映像攻撃の成功には、機材を自立ないように配置するだけでなく、不審に思った第三者によって機材を取り除かれないようにすることが求められる。近年、プロジェクトは小型なものが登場しているため、映像投影機材のステルス性は高まっている。また、自律飛行警備ドローンによる巡回シナリオでは、その場所には巡回を行う警備ドローンのみが存在するため、攻撃の実行が第三者に目撃される可能性はさらに低いと考えられる。

7.2 防御手法

本研究の結果、敵対的映像によって自己位置推定に誤認識が生じることが明らかになり、これを軽減する防御手法の必要性が示唆された。本課題に対応するために、追加センサの搭載が有効であると考えられる。具体的には IMU ならびに LiDAR センサをドローンに搭載することで、機体の移動情報等を推定することができるため、攻撃による影響を低減することが可能である。また、カメラへの入力映像から動的物体を除去することで、敵対的映像の影響を無効化することが可能である。他センサの搭載あるいは動的物体の除去という防御手法の提案と評価は、今後の課題である。

7.3 研究倫理

本稿が指摘する脅威は、特定の製品の脆弱性ではなく、オープンソースの vSLAM システムである ORB SLAM3 に起因するものである。しかしながら、将来的に実社会に及ぼしうる影響を慎重に考慮し、ORB SLAM3 を用いた自律飛行を実装していると考えられる自律飛行ドローンメーカー等に対して責任ある情報開示のプロセスを進めている。また、実世界での実験は、大学構内の管理された環境において安全に配慮して実施している。

8. 結論

本研究は、vSLAM の自己位置推定を意図的に任意の方向に誘導する攻撃手法である敵対的映像攻撃を提案し、その有効性を検証した。敵対的映像攻撃は屋内環境に設置した敵対的映像により、vSLAM の自己位置推定に誤認識を引き起こす手法である。カメラ単体評価では、vSLAM の代表的なアルゴリズムである ORB SLAM3 の自己位置推定に最大で約 252 m の誤差が生じることを示した。また、シミュレーション上及び商用ドローンを活用した実世界上での E2E 評価では、敵対的映像の利用により自律飛行ドローンが任意の方向に誘導され、墜落する可能性があることを示した。本研究により vSLAM ベースの自律飛行システムにおける脆弱性が明らかとなったが、今後はこの脅威に対する実用的な防御手法の開発が急務である。具体的には、IMU や LiDAR などのセンサーを活用したマルチセ

ンサフュージョン環境における攻撃耐性の検証、本研究で提案した動的物体検出・除去アルゴリズムの実装、さらに ORB SLAM3 以外の vSLAM システムや異なる制御アルゴリズムを採用したドローンへの攻撃影響の包括的な評価が重要な課題となる。

謝辞 本研究の一部は JSPS 科研費 22H00519, JST CREST JPMJCR23M4 の助成を受けたものです。

参考文献

- [1] Skydio. Skydio 2+. <https://www.skydio.com/skydio-2-plus-enterprise>.
- [2] Baodong Chen, Wei Wang, Pascal Sikorski, and Ting Zhu. Adversary is on the Road: Attacks on Visual SLAM using Unnoticeable Adversarial Patch. In *USENIX Security 2024*, pp. 905–922, 2024.
- [3] Ben Nassi, Yisroel Mirsky, Dudi Nassi, Raz Ben-Netanel, Oleg Drokin, and Yuval Elovici. Phantom of the ADAS: Securing Advanced Driver-Assistance Systems from Split-Second Phantom Attacks. In *Proc. ACM SIGSAC CCS*, pp. 293–308, 2020.
- [4] Yaniv Nemcovsky, Matan Jacoby, Alex M. Bronstein, and Chaim Baskin. Physical passive patch adversarial attacks on visual odometry systems. In *Proc. ACCV*, 2022.
- [5] Y. Wang, X. Yao, X. Liu, X. Li, P. Hao, and T. Zhu. I can see the light: Attacks on autonomous vehicles using invisible lights. In *Proc. ACM CCS*, 2021.
- [6] John Doe, Jane Smith, and Robert Brown. Perceptual aliasing++: Adversarial attack for visual slam front-end and back-end. *IEEE Robotics and Automation Letters*, Vol. 7, No. 2, pp. 1–1, April 2022.
- [7] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim. Rocking drones with intentional sound noise on gyroscopic sensors. In *Proc. 24th USENIX*, 2015.
- [8] Z. Wang, K. Wang, B. Yang, S. Li, and A. Pan. Sonic gun to smart devices: Your devices lose control under ultrasound/sound. In *Black Hat USA*, 2017.
- [9] D. Davidson, H. Wu, R. Jellinek, T. Ristenpart, and V. Singh. Controlling uavs with sensor input spoofing attacks. In *WOOT*, 2016.
- [10] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. In *Proceedings of the IEEE Transactions on Robotics*, Vol. 37, pp. 1874–1890, 2021.
- [11] U.S. Government Accountability Office. Drone operations. <https://www.gao.gov/drone-operations>.
- [12] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *ECCV*, pp. 834–849, 2014.
- [13] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *ISMAR*, pp. 225–234. IEEE International Symposium on Mixed and Augmented Reality (IEEE), 2007.
- [14] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. In *CVPR*, pp. 1052–1067, 2007.
- [15] Ryze Robotics. Tello. <https://www.ryzerobotics.com/jp/tello>.
- [16] Ouster. Os1: High-res mid-range lidar sensor for automation & security. <https://ouster.com/ja/products/hardware/os1-lidar-sensor>.