

チャッドローン：LLMによる画像認識を用いた自律飛行ドローンシステムの開発と実験

森崎 一步^{1,a)} 湯村 翼^{1,b)}

概要：ドローンはさまざまな分野で活用され、操作が不要な自律飛行技術の開発も進む。GNSS が利用できない屋内環境で自律飛行を実現するためにドローン撮影画像の分析が用いられるが、モデルの適用には限界があり、状況に応じた適切な制御にはプログラムやモデルの開発にコストがかかる。そこで本研究では、大規模言語モデル (Large Language Model : LLM) の画像認識技術を用いてドローンが自律飛行を行うチャッドローンを提案する。チャッドローンでは、画像処理のためのモデルやプログラムを用意するのではなく、LLM に送るプロンプトを用意する。ドローンが撮影した画像とプロンプトを LLM に送り、LLM の応答をもとにして状況に合わせた飛行を行う。小型ドローン Tello と ChatGPT を用いて、チャッドローンのプロトタイプ開発を行った。チャッドローンの有効性を確認するために、予備実験と本実験を行った。プロンプトの記述を詳細にすることでより意図通りの自律飛行を実施させることができるが、プロンプトが長くなるため応答時間も長くなるという相反関係があることが明らかとなった。

キーワード：ドローン, 大規模言語モデル, プロンプトエンジニアリング

1. はじめに

ドローンは災害、物流、エンタテインメントなど、さまざまな分野で活用される。ドローンを有効に活用するためには、熟練した操縦士が必要となる。特に複雑な環境での飛行には高度な操縦技術が求められ、操縦士がいない状況ではドローンの運用が制限される。そのため、自律飛行技術の開発が進められる。屋外では GNSS を用いたナビゲーションが一般的であるが、屋内環境では GNSS 信号が利用できない。それを解決する手法のひとつが、画像処理技術を活用した飛行制御である。ドローンが撮影する映像データをリアルタイムで解析し、障害物の回避や飛行経路の最適化を自律的に行う。ただし、飛行制御を行うための画像処理プログラムや画像処理のための学習モデルは事前に作成しておく必要があり、現地の状況に応じて制御方法を変更するには大きな開発コストがかかる。モデルが状況に合致しない場合には、モデルの再構築が必要となる場合もある。

そこで本研究では、より汎用性の高い大規模言語モデル



図 1: チャッドローン概要

(Large Language Model : LLM) の画像認識技術を用いて自律飛行を行うドローンシステムチャッドローン (図 1) を開発する。チャッドローンでは、画像処理のためのモデルやプログラムを用意するのではなく、LLM に送るプロンプトを用意する。ドローンが撮影した画像とプロンプトを LLM に送り、状況に合わせた飛行を行う。本論文では、システム設計、プロンプト設計、屋内で実施した飛行実験の結果について記述する。

2. 関連研究

LLM を使用したドローン制御の研究がいくつか存在する。池山ら [1] は、ドローンの広域景観撮影に LLM を用いた。自律飛行のウェイポイントの設定に LLM を用い、専門知識のないユーザでもドローンを活用できるようにし

¹ 北海道情報大学
Hokkaido Information University
a) s2121155@s.do-johodai.ac.jp
b) yumu@yumulab.org

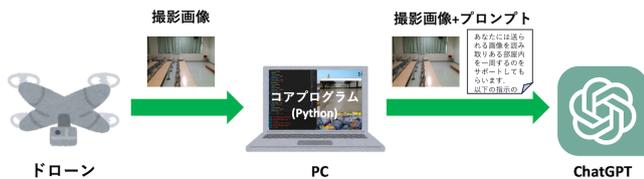


図 2: 画像とプロンプトの送信

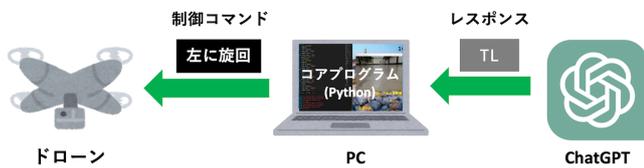


図 3: レスポンスとドローン制御コマンドの送信

た。この研究では屋外の自律飛行を対象としたが、チャットドローンは屋内での自律飛行を対象とする。

Vemprala *et al.*[2] は、ドローンやロボットアーム等のロボティクスへの LLM の適用に関して、設計原則とデモンストレーションが示された。ドローンのデモンストレーションでは、人間が都度指示を行う対話的インタフェースが用いられた。チャットドローンでは、ルールを記述したプロンプトをあらかじめ設定しておき、途中で人間が関わらない形の自律飛行を実施する。

3. チャットドローン

3.1 概要

チャットドローンでは、LLM として ChatGPT[3] を用いる。ドローンの制御と ChatGPT との対話はコアプログラムが担う。コアプログラムは PC 上で動作する。コアプログラムは、ドローンから撮像画像を受信し、画像にプロンプトを付与して ChatGPT に送信する (図 2)。プロンプトは、画像をもとに現在の状況を判断するよう指示する内容である。プロンプト内では次の行動の選択肢を提示する。ChatGPT は、撮像画像をもとに状況判断を行い、現在の状況を選択肢から選んでレスポンスとして返す。コアプログラムはレスポンスを受信し、これ元に、次の行動の制御コマンドをドローンへ送信する (図 3)。これを繰り返すことで、ドローンの自律飛行を実現する。

3.2 実装

本研究では、チャットドローンのプロトタイプ実装にあたり、ドローンに Ryze Tech 社の Tello[4] を用いた。Tello は、外部からの制御コマンドを受け付ける WebAPI を有する。また、搭載カメラで撮影した画像を外部に送信することもできる。これらは、Tello がアクセスポイントとなる形で Wi-Fi 経由でアクセスできる。Tello への制御コマンドの送信は、PC で稼働するコアプログラムが担う。コアプログラムは Python で実装した。

コアプログラムは、ChatGPT の利用のための OpenAI



図 4: システム構成

あなたには送られる画像を読み取りある部屋内を一周するのをサポートしてもらいます。
 以下の指示の内画像の状況にあてはめ、丁度当てはまるものを選び答えを出力して下さい

指示一：右側に壁や壁のようなもの（窓やドアでも可）がある状態で前方に二メートル以上の空間がある場合は M と出力して下さい

指示二：右側に壁や壁のようなもの（窓やドアでも可）がある状態で前方に二メートル以上の空間がない場合は TL と出力して下さい

図 5: プロンプト Ver.1

API のアクセスも担う。コアプログラムが稼働する PC は Tello との通信に WiFi を用いるが、これにはインターネット到達性がない。そのため、OpenAI API へのアクセスのために、イーサネット接続回線を利用する。実験時には、イーサネット接続が可能なホームルーターを用いた (図 4)。

コアプログラム上でチャットドローンの開始を指示すると、ドローンは飛翔する。ドローンが指定された高度に達すると、写真の撮影を行う。コアプログラムは撮像画像を受信し、プロンプトとともに ChatGPT に送信する。ChatGPT は送信されてきた画像を解析し、現在の状況をレスポンスとしてコアプログラムに送信する。ChatGPT からのレスポンスを受信したコアプログラムは、レスポンスの内容に応じてドローンへ制御コマンドを送信する。

3.3 プロンプト

ChatGPT に送信するプロンプトは、本システムにおける非常に重要な要素である。コアプログラムでの処理が可能となるよう、ChatGPT のレスポンスに特定の文字列を含める必要があるが、そのためにはプロンプトの工夫が必要となる。プロンプトの構成に以下のルールを設けた。

- はじめに概要を記述
- 画像を読み取り、当てはまる状況を選択肢から選択して返答する旨を記述
- 状況の選択肢を記述
- どの選択肢も当てはまらない場合の指示を記述 (Ver.1.1 から追加)

これらのルールに基づき、屋内の矩形の部屋において、壁に沿って前進し、衝突しそうな場合は進路を左に転回するというプロンプトを作成した。プロンプトは、後述する実証実験を経て作成した。Ver.1 と Ver.2 をそれぞれ図 5 と図 6 に示す。

あなたには送られる画像を読み取りある部屋内を一周するのをサポートしてもらいます。なので以下の指示を画像の状況にあてはめ、丁度あてはあるものを選び答えを出力してください。

指示一：右側に壁や壁のようなもの（窓やカーテンやドアでも可）がある状態で前方に二メートル以上の空間がある場合は M と出力してください。

指示二：右側に壁や壁のようなもの（窓やカーテンやドアでも可）がある状態で前方に三メートル以上の空間がある場合は MMM と出力してください。

指示三：右側に壁や壁のようなもの（窓やカーテンやドアでも可）がある状態で前方に五メートル以上の空間がある場合は MM と出力してください。

指示四：右側に壁や壁のようなもの（窓やカーテンやドアでも可）がある状態で前方に二メートル以上の空間がない場合は TL と出力してください。

指示五：左右に壁や壁のようなもの（窓やカーテンやドアでも可）がなく前方に二メートル以上の空間がある場合は M と出力してください。

指示六：左右に壁や壁のようなもの（窓やカーテンやドアでも可）がなく前方に五メートル以上の空間がある場合は MM と出力してください。

指示七：左右に壁や壁のようなもの（窓やカーテンやドアでも可）がなく前方に二メートル以上の空間がない場合は TL と出力してください。

また画像から判断できない場合は TL と出力してください

図 6: プロンプト Ver.2



図 7: 予備実験を実施した大教室



図 8: 予備実験でのドローン撮影画像

4. 実証実験

4.1 概要

チャドローンの提案手法の有効性を確認するため、実証実験を行った。実験は、予備実験と本実験の2度実施した。いずれの実験も、屋内の矩形の部屋において部屋のなるべく外側の壁沿いを周回するという飛行を想定して実施した。実証実験において確認したい項目を以下のものに絞った。

- (1) ChatGPT が意図通りのレスポンスを作成するか
- (2) LLM によってドローンを自律飛行させることが可能か
- (3) ドローンを意図通りに動かすことが可能か
- (4) プロンプトの内容を変更することでふるまいがどのように変化するか

4.2 予備実験

予備実験は、大学内の大教室(図7)にて実施した。Ver.1のプロンプトを使用して実験を行い、撮影した画像(図8)を用いて ChatGPT が分析し、意図通りに動作させることが可能であることを確認した。実験項目(1)~(3)が可能であることを確認した。一方、使用したドローンの Tello が軽量なため、空調の換気の風が移動の妨げとなった。そのため、場所を変えて本実験を実施した。

4.3 本実験

本実験は、ゼミナール用の小教室(図9)で実施した。教室の大きさは4.7m × 10.1mである。教室には机、椅子、



図 9: 本実験を実施した小教室

キャビネットがあり、実験実施時も配置したままとした。本実験では、Ver.1 と Ver.2 の2種類のプロンプトを用い、それらの比較もおこなった。

まず、プロンプト Ver.1 を用いて実験を実施した。ChatGPT の動作が正常、つまり特定文字列を含むレスポンスを返し続ける場合には、想定通り教室内の壁に沿って移動することが確認できた。しかし、ChatGPT が指示外のレスポンスを返すことがまれに発生した。

次に、実験項目(4)の確認のためにプロンプトを修正し、プロンプト Ver.2 を用いて実験を実施した。プロンプト Ver.2 では、「画像から判断できない場合は TL と出力して

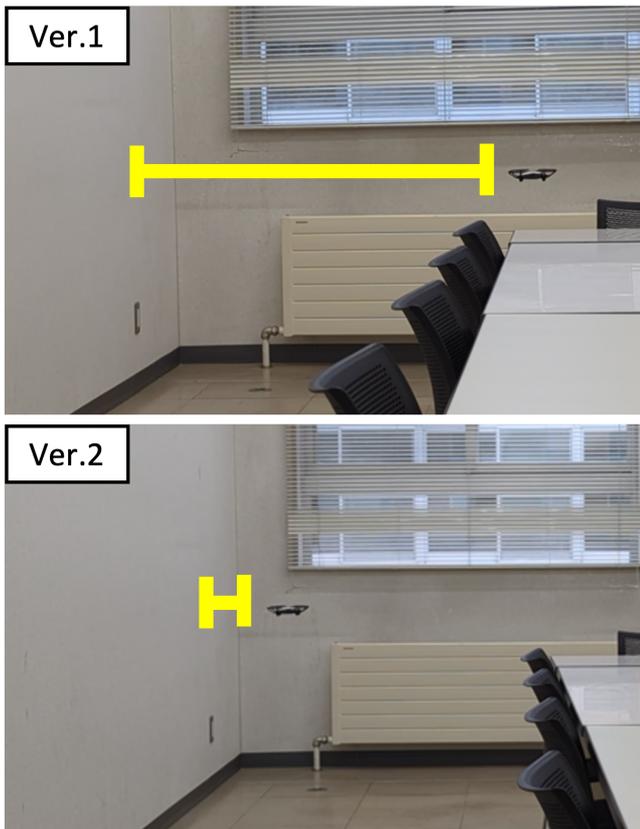


図 10: プロンプト Ver.1 と Ver.2 におけるふるまいの比較。黄色の線は、転回時の壁とドローンのおおよその距離を示す。

ください」という一文を追加することで、例外処理を実施するようにした。

状況判断の選択肢は、Ver.1 では「直進可能」「直進不可(左転回)」の 2 種類のみであったが、Ver.2 では、直進可能な距離に応じて 7 種類を提示した。

図 10 は、プロンプト Ver.1 利用時とプロンプト Ver.2 利用時でドローンの壁との距離を比較したものである。プロンプト Ver.1 では壁との距離が大きく、プロンプト Ver.2 では壁との距離が小さい。プロンプトの内容を変えることで、ユーザの意図をより反映させて飛行させることが可能であることがわかった。

5. おわりに

本研究では、操縦士を必要としないドローンの実現を目指し、LLM を活用した自律飛行ドローンシステムチャッドロンを提案した。そのプロトタイプを開発し、実証実験を実施した。実験の結果、LLM を用いてドローンの自律飛行が可能であることを確認した。プロンプトを変更することで、ドローンのふるまいを変えられることも確認した。

チャッドローンにおける、いくつかの課題も明らかとなった。制御コマンドの送信のインターバルに割り込み制御を行う仕組みがないため、急激な環境変化に対応できな

い。プロンプトの記述詳細にすることでより意図を反映させた自律飛行を実施させることができるが、プロンプトが長くなるため、ChatGPT の応答時間も長くなる。Ver.1 と比べて Ver.2 のプロンプトでは、応答時間が 3 倍以上となる場合もあった。

チャッドローンでは、ChatGPT に送信するプロンプトが自律飛行における肝である。プロンプトは自然言語であるため、専門的知識がないユーザでも書き換えることができる。将来的には、ユーザがプロンプトの更新をして、パーソナライズされた自律飛行をデザインするという使い方も考えられる。本研究の成果が、LLM を活用した自律飛行ドローンの可能性を広げ、実用化の一助となることを期待する。

参考文献

- [1] 池山安杜里, 山内翔, 鈴木恵二: 大規模言語モデルによるドローンの広域景観撮影システムの提案, *IEICE Conferences Archives*, The Institute of Electronics, Information and Communication Engineers (2023).
- [2] Vemprala, S. H., Bonatti, R., Bucker, A. and Kapoor, A.: Chatgpt for robotics: Design principles and model abilities, *IEEE Access* (2024).
- [3] OpenAI: ChatGPT, <https://chat.openai.com>.
- [4] RyzeTech: Tello, <https://www.ryzerobotics.com/jp/tello>.