

リモートカメラ制御シナリオのための モデル生成手法

キヤノン株式会社

黒田賢

kurota.masaru@mail.canon

撮影現場の抱える課題

< 背景 >

- ・コスト削減、省人化のニーズが高まっている
- ・IP経由での制御が可能なカメラの導入
⇒従来では人が入れない位置からの撮影も可能

< 課題 >

- ・コントローラ操作には従来と異なる熟練度が必要
- ・複数台のカメラを同時、切り替えて制御が必要

手法・ツールの適用による解決

< 解決策 >

ディレクターやカメラマンの口語的・定性的な指示を、
制御プロトコルへの変換してIP経由で送信し、
カメラ制御を行うシステムを構築する

ユーザ	システム	カメラ
口語的指示 定性的指示	→ プロトコル変換 IP送信	受信データに 基づいて動作

検討フロー

【検討フロー】

仮想環境にて一連の動作を構築・評価してから実機検証へ

仮想環境

評価

実機環境



ローカルLLM

- ・遅延が少ないことが好ましい
- ・一部機能を含む可能性がある

データセット

- ・自然言語 ⇒ CANON XC プロトコル

Fine tuning

- ・日本語学習済みのモデルをベース
- ・独自データセットによる再学習

入出力シミュレータ

- ・プロンプトに相当するテキスト入力
- ・LLMの出力を受信するカメラシミュレータ

実カメラ

- ・非機能
- ・異常動作

※動作確認まで

想定入力において所望の出力となるかを検証・評価

システム概要

【データセット例】

```

"instruction": "Canon XC Protocolに変換してください",
"input": "左30度パンしろ。",
"output": "/-wvhttp-01-/control.cgi?pan=-3000"

```

```

"instruction": "Canon XC Protocolに変換してください",
"input": "高速で上にチルトしろ。",
"output": "/-wvhttp-01-/control.cgi?tilt=up&tilt.speed=10000"

```

プロトコルや撮影用語を理解していない
モデルに対して、誘導が必要となる。
定量的な指示と、定性的な指示を
織り交ぜる形で学習データを形成した

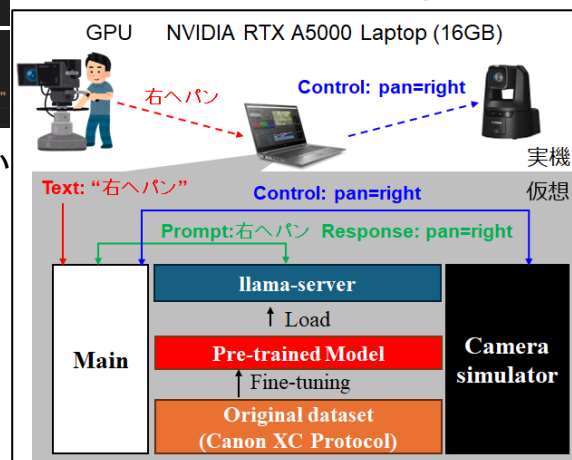
【入出力】



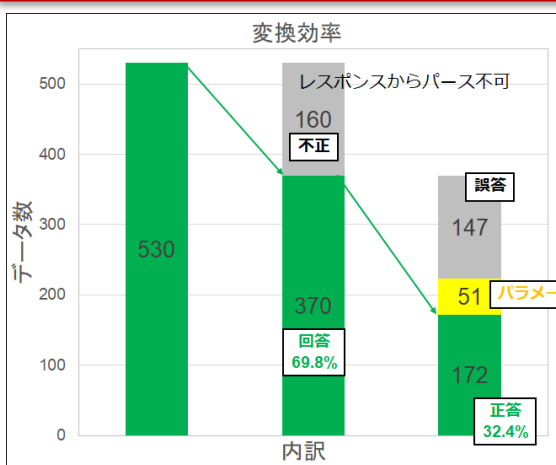
対話ではなくサーバー動作での検証

【全体像】※学習環境を除く

オリジナルのデータセットにて、日本語学習済みのモデルをFine-tuningし、
テキストからプロトコルへの変換をさせた

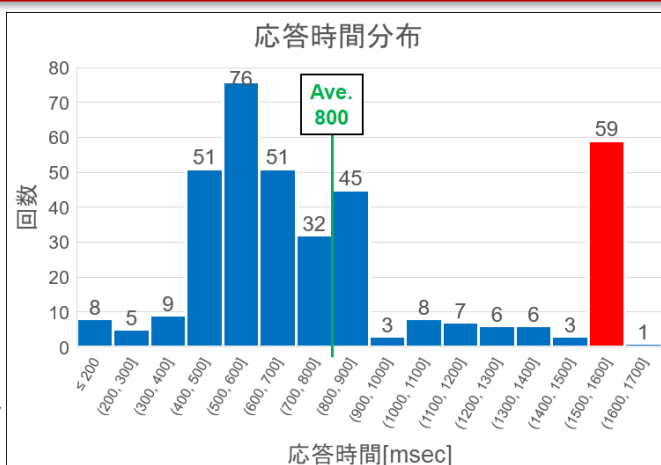


変換精度と非機能



※106個 × 5回 = 530個のデータを変換
プロトコル解釈可能なデータ **69.8%**
期待値通りのデータ **32.4%**

失敗は放送事故なので100%が目標



解釈可能なデータ内での応答
特異的に遅いデータ(1500msec強)があったが
500~900msecがボリュームゾーンである
リアルタイムではなくバックグラウンド制御向き

対話での動作

```

> じわじわ下にチルト
/-wvhttp-01-/control.cgi?tilt=down&tilt.speed=100

> さっと上にチルト
/-wvhttp-01-/control.cgi?tilt=up&tilt.speed=500

> 素早く上にチルト
/-wvhttp-01-/control.cgi?tilt=up&tilt.speed=10000

```

定性的なニュアンスを理解し、パラメータを
変化させることが出来ていることを確認

今後の展望

【システム見直し】

- ・プロトコルを分解し、制御種別はルールベース、LLM+RAGによるパラメータ個別補正による組み合わせ

【モデル改良・軽量化】

- ・精度と非機能を両立するチューニング