ノート

ロボットの高度制御技術の開発

木嶋　祐太*　高橋　靖*　長谷川　直樹*

Development of AI-Enabled Autonomous Robot Control Technologies

KIJIMA Yuta*, TAKAHASHI Yasushi* and HASEGAWA Naoki*

抄　　録

本研究は専門的なプログラム無しで自律的に動作し，製造工程の自動化を実現するロボットの開発を目的としている。本研究では，ロボットに求められるAI技術の調査およびAIエージェントによるロボットの制御，模倣学習によるロボット行動生成モデルの学習について実施した内容を紹介する。

1.　緒　　言

国内製造業は就労人口の減少や高齢化へ対応しながら生産性を向上することが求められている。解決の手段として，画像処理やAIと連携した高度なロボットの制御技術が注目されている。

ロボットを制御するためにはプログラムを作成する必要があり，現在，ノーコードツールなどによりそのハードルは下がっているが，最新のAI技術を用いることで，さらに運用が容易になり，導入障壁を低くすることが期待できる。本研究は専門的なプログラム無しで自律的に動作し，製造工程の自動化を実現するロボットの高度制御技術について検討した。

2.　AI技術の調査

2.1　生成AI，AIエージェント

近年急速に普及している生成AIやAIエージェントは，ロボット制御において，環境理解，作業手順の自動生成，状況に応じた行動選択を担う知能層として機能することで，従来の制御則やルールベースを補完し，人の意図を柔軟に反映した自律的なロボット動作の実現につながることが期待されている。VLM(Vision Language Model)と呼ばれる画像と言語に対応したモデルによるロボットの周辺環境認識，複数のロボット群の統括へのAIエージェントの応用，通信手段を持たないロボット同士の連携への音声の活用，といった事例が報告されている。

2.2　基盤モデル，物理AI

基盤モデルは大量の画像・言語・行動データから汎用的な知識や判断能力を学習し，環境変化への柔軟な対応を可能にしたモデルであり，物理AIは，センサで観測された実世界の状態を入力とし，物理法則・機構制約・時間連続性を前提に，行動や制御量を生成するモデルである。これらを組み合わせることで，教示や再設計の負担を減らした高度なロボットの実現が期待されている。

3.　ロボットの高度制御技術の開発

3.1　AIエージェントによるロボットの制御

AIエージェントは，目標を与えると，状況を観測・判断し，必要な行動を自律的に選択・実行するAIのことである。対話，計画，ツール操作，他システムとの連携を通じて，単なる回答に留まらず，タスクを継続的に遂行する点が特徴である。本研究ではAIエージェントがロボットを制御するシステムを試作した。

ロボットを動作させるツール，画像を取得するツール，画像と物体の名前から物体の位置を取得するツールの3つを定義した。AIエージェントはこれらのツールを使用してタスクを実行

* 技術統括センター

する。例えば「キューブを持ち上げる」というタスクが与えられたら，AI エージェントは，各ツールを使用し表 1 の順番で実行する。対話によりタスクを遂行することもあり，複数の物体が見つかった場合にはどの物体を持ち上げるかを指示者に聞く。

図 1 に示す RGB-D カメラとロボットからなるデモ装置を作成した。RGB-D カメラは Intel 社製 RealSense D405i を，ロボットは RT Corporation 製 CRANE+ V2 を使用した。デモ装置では RGB-D カメラとロボットの座標系を合わせるため AR マーカーを使用した。図 1 に示すキューブ，テープ，消しゴムを検出する物体とした。物体をランダムに配置して，表 2 に示すタスクをそれぞれ 12 回ずつ実施した。X のみのタスクは 3 種類の物体を 4 回ずつ，X と Y があるタスクは 6 通りの組み合わせを 2 回ずつである。表 2 に結果を示すが，失敗はすべて物体の位置を取得するツールのミスが原因であった。

重さを測るタスクではデジタルのキッチンはかりを用いた。システムは画像を取得できるため，物体を乗せた後に撮影した画像を AI エージェントが解析し数値を読み取ることができる。ネットワークに接続できる機器でなくても結果を画像で読める機器はシステムに活用できることがわかった。

3.2 模倣学習によるロボット行動生成モデルの学習

物理 AI に活用されることが想定されるロボット行動生成モデルである ACT（Action Chunking Transformer）[1]および GR00T[2]を模倣学習により学習し，動作実験を実施した。模倣学習とは，人の操作データをもとにロボットの制御を学習する手法である。

ACT および GR00T は，画像，関節角，言語プロンプト（言語プロンプトは GR00T のみ）を入力とし，複数ステップ分の関節指令列を出力するロボット行動生成モデルである。ACT は，行動を一定区間ごとにまとめて生成するチャンク構造を採用し，過去の流れを考慮しながら行動を生成する。さらに潜在変数により行動のばらつきも表現できるため，予測で生じる誤差の累積を抑え，複雑で揺らぎのある操作タスクに適しているという特徴を持つ。GR00T は，大規模な視覚と言語データを事前学習した基盤モデルを基に，ロボットの観測情報から直接行動を生成するモデルである。多様な環境やタスクに対して汎用的に適応するために，少量の追加データによる微調整が想定されている。

図 2 に示すデモ装置を作成し，ロボットがキューブを収納場所に運ぶタスクを実験した。画像を撮影するためのカメラはカメラ 1 とカメラ

表 1 「キューブを持ち上げる」の実行手順

1	ロボットを動作させるツールで，ロボットを撮影の邪魔にならない姿勢に移動。
2	画像を取得するツールで，環境を撮影。
3	物体の位置を取得するツールで，画像からキューブの位置を取得。
4	ロボットを動作させるツールで，ロボットにキューブを把持。
5	ロボットを動作させるツールで，キューブを持ち上げ。
6	「キューブを持ち上げました」と回答。

図 1 AI エージェントデモ装置

表 2 AI エージェントでの実施タスクと結果

タスク	結果（成功率）
Xを持ち上げる	11/12回成功（92%）
XをYの上に乗せる	9/12回成功（75%）
Xの重さを測る	8/12回成功（67%）

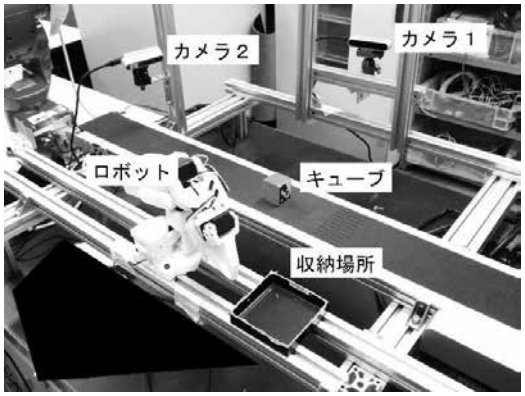

図 2 ロボット行動生成モデルのデモ装置

2 の位置に配置した。カメラ 1 は Intel 社製 RealSense D415i を，カメラ 2 は Intel 社製 RealSense D435i を使用した。ロボットは LeRobot[3]プロジェクトが公開しているオープンソースロボット SO-101 を使用した。人間がロボットを操作して，キューブを収納場所に運ぶ動作を 250 回行い学習データを作成した。学習データの作成条件を表 3 に示す。

図 3 に示す●位置の 9 か所にキューブを配置して，カメラ 1 に対して正対および斜めの 2 条件でキューブを収納場所に運ぶテストを実施した。テストの結果を表 4 に示す。タスクは極めて単純であるが，どちらのモデルも成功率が高いとは言えなかった。

4. 結　　言

(1) AI 技術を調査し，AI エージェント，物理 AI といった技術のロボット制御への応用が期待されていることがわかった。

(2) AI エージェントにより，ツールを定義することで，人が与えたタスクを解釈し，自律的にロボットが動作するシステムが開発できることを確認した。

(3) プログラミングをせずに，ロボット行動生成モデルを模倣学習し，ロボットを自律的に動作させた。多くの学習データが必要な割に，タスクの成功率は高くなかった。

(4) 物理 AI の活用を見据えて，シミュレーターによるデータ生成などの手段でロボット行動生成モデルのタスク成功率を向上する研究を今後実施する。

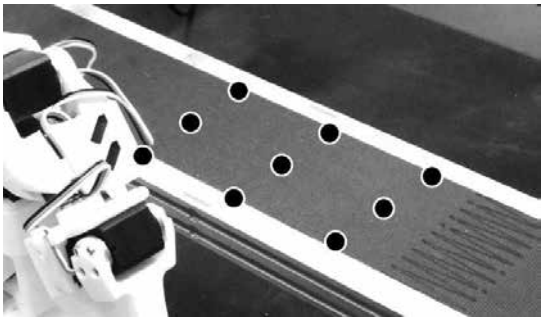

図 3 テストのキューブの配置位置

表 3　学習データの作成条件

データ数	250
画像サイズ	320×240
サンプリングレート	10 Hz
ロボットの関節角数 (グリッパも含む)	6
GR00T の学習に用いたテキスト指示	Pick up the yellow cube and move it to the specified fixed location on the workspace.

表 4　ロボット行動生成モデルのテスト結果

ロボット行動生成モデル	結果（成功率）
ACT	11/18回成功（61%）
GR00T	9/18回成功（50%）

参考文献

1) tonyzhaozh act, https://github.com/tonyzhaozh/act, (閲覧 2026-2-13).

2) NVIDIA Isaac-GR00T, https://github.com/NVIDIA/Isaac-GR00T, (閲覧 2026-2-13).

3) LeRobot, https://github.com/huggingface/lerobot, (閲覧 2026-2-13).