

# 2025年度第7回ロボフレ委員会講演内容 ～事業化を目的とした生成AI利用の産業用ロボット運動制御～

講師：立命館大学 川村貞夫、記録：RRI 西垣戸貴臣

概要：本稿は、労働人口減少を背景とする産業現場の自動化需要に対し、生成AIを活用した産業用ロボット動作自動生成手法を提案するものである。従来の教示再生方式は高い信頼性を有する一方、熟練SIerへの依存度が高く、環境変動への適応性に乏しいという課題を抱えている。近年注目される基盤モデルによるエンドツーエンド制御は高い汎用性が期待されるが、膨大な学習データを必要とし、産業用途で要求される精度・実時間性・信頼性を1年程度の短い研究開発期間で達成することは困難と予想される。

本研究では、生成AIをロボットシステムの中核に据えるのではなく、限定的な機能要素として位置づけ、AI依存度を抑制した分割型アーキテクチャを採用する。具体的には、対象物認識および作業・運動認識にのみ生成AIを適用し、環境認識および身体認識には決定論的手法を用いることで、再現性と安定性を確保する。また、絶対座標系の高精度同定を不要とし、相対誤差の収束のみで高精度動作を実現する独自のビジュアルフィードバック制御を組み合わせることで、産業用途に求められる精度を達成している。

適用例としてピックアンドプレイス作業を対象に、低コストなハードウェア構成で短時間の教示により実用的なロボット動作が生成可能であることを実証した。さらに、個別作業単位で構築した知能を段階的に統合する「Individual Task Intelligence Integration」という発展戦略を示し、短期実装と将来的拡張性を両立する産業用ロボット知能化の方向性を提示している。

## 1. はじめに

現在、多くの産業用ロボットの動きは、ティーチングプレイバックと呼ばれる手法を用いて生成されている。これは、人間がロボットに動きを教え、ロボットはそれを繰り返すという手法である。一方、近年基盤モデルという手法が盛んに研究されている。これは、最上位の自然言語による指令を用いてロボットの動作を自動生成するものであり、画期的な手法として注目されている。しかしながらこの手法は膨大な学習データを必要としており、かつ、産業用ロボットに必要とされる精度の点で課題が残されている。

本稿は、産業用ロボットに活用可能な精度を有し、かつ、1年内に実現可能な、生成AIを用いたロボット動作の自動生成手法を提案するものである。第2章では背景となる人口減少を解説する。第3章では、提案手法の技術的なバックグラウンドとなるビジュアルフィードバック技術について説明する。第4章では、課題解決の考え方を示す。第5章で具体的な手法を提案し、第6章で、実証の結果を述べる。第7章はまとめである。

## 2. 労働人口の減少

日本では人口減少が進む一方、就労人口は高齢者・女性の就労拡大により直近10年ほどは微増で維持されてきた。しかし今後10年はその支えが限界に近づき、社会を維持するために「大量の外国人労働

者受け入れ」か「自動化技術（ロボット等）による代替」かの正念場にある、という危機感が背景に置かれる（図1）。特に社会を支えるエッセンシャルワークの維持が重要で、業務をデスクワーク／ノンデスクワークに分けると、前者はDX・AIの恩恵で生産性向上が進む一方、後者（現場作業）は、物体を「ここから持ってきてここへ置く」だけでも難易度が高く、ロボット（RX）が本来担うべき領域だが、依然として課題が多い。講演者らはこの領域で、現場の自動化を少しでも前進させるためにロボットシステム開発を進めている。

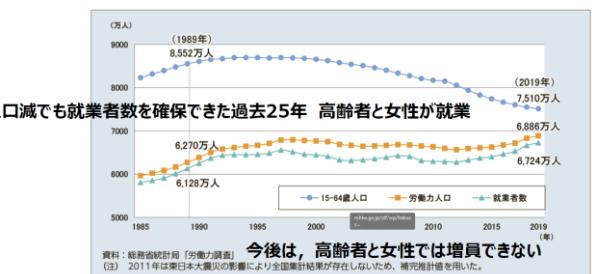


図1 労働人口問題

## 3. ビジュアルフィードバック

ここで説明する技術は、従来のビジョン誘導と異なる原理のビジュアルフィードバックである。通常はカメラキャリブレーション等により、ロボット先端位置や目標位置をある座標系で高精度に同定し、

幾何モデル（運動学）や動力学モデルを前提に制御を設計する。しかし提案方式では、ロボット先端と目標位置の「相対誤差」だけが分かればよく、先端がどの座標系でどこにいるかを絶対値として知る必要がない。したがって提案制御部では、運動学・逆運動学・ダイナミクス・逆ダイナミクスを用いずに実現できる。産業ロボットコントローラ内部の既存機能として一部の逆運動学等は使われるが、提案する外付け制御設計としては不要である。このため、レンズ歪みを含むカメラの「物差し」が完全に正しくなくても、誤差がゼロへ収束すれば目的が達成される。床の歪みやカメラ設置角度のズレなど、現場で避けにくい不整合が残っても成立しやすいことが利点として挙げられる。



図2 従来のビジュアルフィードバック制御

- ・座標系は制御には不要
- ・対象物とロボットハンドの相対位置誤差のみの情報から制御

#### 新しいロボットSI法を開発

立命館大学川村研2010年から研究開始

##### 特徴

- ・DynamicsもKinematicsも未知でよい
- ・座標系設定がいい加減でも、  
レンズ歪みがあっても高精度に制御可能



図3 講演者らのビジュアルフィードバック制御

コスト・時間と精度の観点では、従来方式より簡便で、より高精度領域まで到達できる。具体的には、繰り返し精度相当の領域までビジュアルフィードバックが可能で、約 20μm 程度の精度で達成できる。一方、従来は人間が努力してキャリブレーションしても 0.5mm 程度が限界という経験則がある。近年 AI でキャリブレーションを徹底した中国企業で 0.2mm 程度という話も聞くが、オーダーが一段違い、本手法に優位性がある。例えば本手法を使えば、安価な Web カメラ相当のカメラを先端に装着し、針と糸で糸通しを行うことができる。糸の中心位置決めは座標系同定が難しいが、先端だけを追えばよい原理により、床の歪みやカメラ角度ズレがあっても実行できる。

#### 4. 生成 AI を用いたロボット運動制御の課題解決

本報告では、生成 AI 及び、前章で説明したビジュアルフィードバック技術を活用し、SI 作業負担を可能な限り削減して、最終的には「SIer がほぼ不要で、ユーザーが使いこなせる」水準に近づける手法を提案する。

ただし時間軸は遠未来ではなく、喫緊の社会課題に対して「一年後に利用可能」な技術を作るという現実的な目標を掲げる。その観点から、基盤モデルには①身体性（身体依存で学習コストが膨大、身体が変わると再学習が必要）、②汎用性（境界が曖昧なまま汎用を追うと効率が上がりにくい）、③実時間性（計算負荷が大きく高速運動に不向きに見える）という三課題がある。また立場は AI 研究者ではなく SI の研究開発であり、生成 AI はロボットの“パーツ”として扱うべきだというスタンスを取る。

加えて、ロボティクスの根源的難しさを「知能・身体・環境（対象物）」が複雑に絡み合う“三体問題”に例える（図4）。相互依存が強く、どれかを決めようとしても別の要因に引きずられるため設計が難しい。

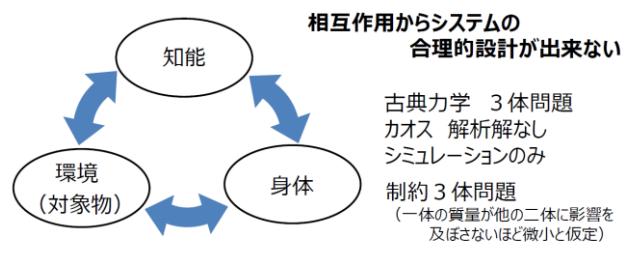


図4 ロボティクスの3体問題

制約付き三体問題のように、解ける形へ落とすには制約導入が重要である。本提案では制約を明確に置く。身体は産業ロボットアーム+市販ハンドに限定し、信頼性を確保する。認識のための探索的接触（子どもの学習のような触って学ぶ過程）は産業ロボットが苦手なので排除する。知能は汎用を求めず、まず現場ニーズの高いピック&プレースに限定する。環境も水平面から水平面、障害物なしなど単純条件から始めるが、扱う対象は現場ニーズのある工業製品で、1mm 級の微細物も視野に入れる（図5）。

## 身体を制約

### 方針①産業用ロボットアームと市販ハンドに限定

- ・高速高精度
- ・高信頼性
- ・高耐久性
- ・高剛性
- ・不要不測の接触の回避必要

### 方針②認識のための運動は可能な限り排除

- ・産業用ロボットアームとハンドは接触が苦手

## 知能を制約

### 方針③ピックアンドプレイス作業に限定

- ・製造業などで広く要求される作業
- ・専用機間の移載作業で人手不足
- ・現状の作業スピード以上要求

## 環境（対象物）を制約

### 設計方針④水平面から水平面への障害物なしの移載作業

- ・環境限定

### 設計方針⑤現在現場でニーズのある対象物

- ・1 mm程度の対象物を含む

図5 ロボティクス3体問題解決方針

## 5. 具体的な手法の提案

従来の産業ロボット分野では、教示再生方式が広く用いられてきた。この方式では、ユーザーが自然言語や画像等を通じて作業要求を SIer（人間）に伝達し、SIer が要件定義、作業計画、運動制御設計を行い、既存のロボットコントローラを用いてシステムを構築する。教示再生方式は、長年にわたり産業現場で高い実績を有している。しかしながら、本方式は本質的に人間による操縦型システムであり、ロボットの自律性は限定的である。作業は人間のノウハウを別の時間軸で再現する形で実行されるため、環境変動や不確実性に対する適応能力に乏しい。また、SIer およびティーチングマンと呼ばれる熟練技術者への依存度が高く、これら人材の高齢化および減少が進行し、その担い手が減少しており、従来手法の継続的運用には構造的な制約が生じている。

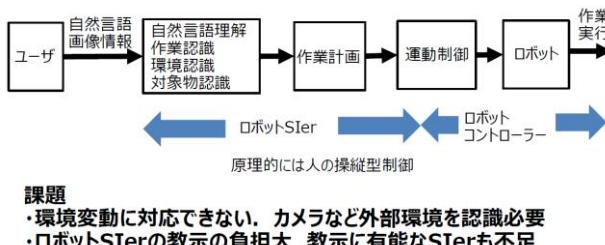
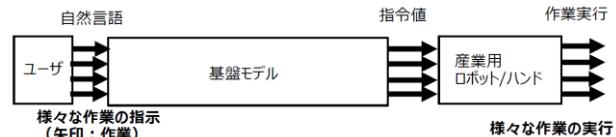


図6 産業用ロボットの教示再生方法

これらの背景を踏まえ、近年では基盤モデルを用いたロボット運動制御が注目されている。自然言語、画像、動画等を入力とし、基盤モデルが作業内容を理解した上で、エンドツーエンドでロボットを制御する構成は、高い汎用性を有する可能性がある。しかし、産業ロボットを対象として、短期間、特に1年以内での現場実装を目指し、かつ実用上要求され

る速度、精度、信頼性を同時に満たすことは現時点では困難であり、技術的・運用的課題が顕在化している。



(1) 汎用性 (2) 身体性 (3) 実時間性の問題が発生

図7 基盤モデルによる運動制御

そこで著者らは、AIへの依存度を意図的に抑制した分割型アーキテクチャを採用した。本方式では、ユーザー入力を自然言語および画像として受け取り、機能を「対象物認識」「作業・運動認識」「環境認識」「身体（ロボット）認識」の4要素に分解する。このうち、生成AIは対象物認識および作業・運動認識に限定して適用し、環境認識および身体認識については従来技術に基づく決定論的手法を用いることで、システム全体の安定性および再現性の確保を図っている（図8）。

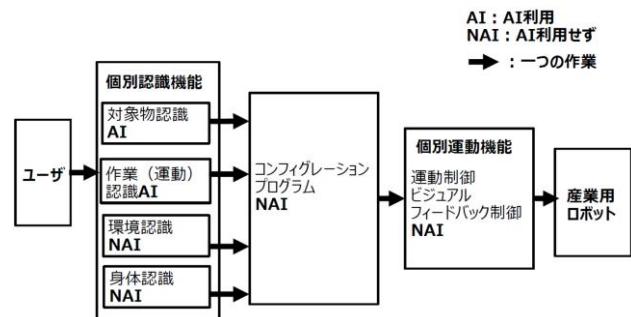


図8 提案システムの構成

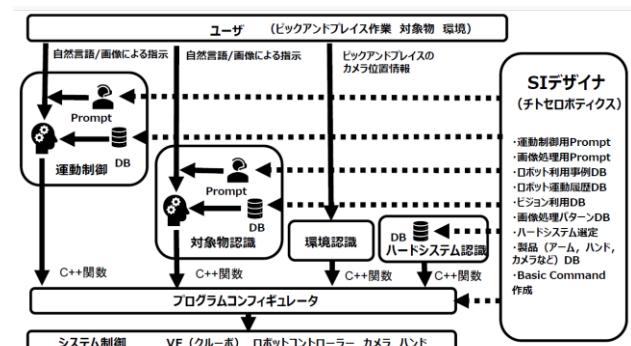


図9 ピック＆プレースを対象とした構成例

生成AIは、最終的にC言語またはC++によるプログラムコード断片を出力する。これらのコードは、非AIベースで構築されたプログラムコンフィギュレ

ータによって統合され、実行可能なロボット制御プログラムとして構成される。生成された制御プログラムは、ビジュアルフィードバックを備えた産業用ロボット制御システムに入力され、実機ロボットの動作を実現する（図9）。

本稿では、具体的な適用例としてピックアンドプレイス作業を対象としたシステム構成を示す。クラウド上の生成AIを用い、自然言語および画像入力から作業内容および対象物を認識させるとともに、著者らが設計したプロンプトおよび作業・運動データベースを活用する。環境認識については、当該作業において必要性が限定的であるため、現段階ではAIを用いていないが、将来的な拡張は可能である。身体認識については、特定の産業ロボットおよび市販ハンドを前提とした構成とし、既存データベースから対応関数を選択する方式を採用している。

これらの要素を統合する役割を担うのがSIデザイナーであり、プロンプト設計およびデータベース整備を行っている。将来的には、ユーザー自身がこれらの工程を担い、SIデザイナーの関与を最小化する運用形態も想定される。

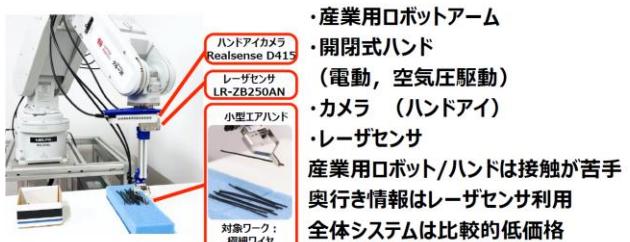


図10 ピック＆プレース作業用ロボットプラットフォーム

ハードウェア構成としては、産業用ロボット、市販ハンド、RealSense カメラ、レーザー距離センサを組み合わせ、高精度かつリアルタイム性の高い認識を実現している（図10）。微細対象物を扱うピックアンドプレイス作業にも対応可能であり、本構成は特定作業における標準的プラットフォームの一例を示すものである。今後は、作業内容に応じて多様なハードウェアプラットフォームが構築されると考えられ、本研究はその技術的基盤を提供するものである。

## 6. ロボットに実装した例

本章では、比較的低コストなハードウェア構成を用いた産業用ロボット向けピックアンドプレイス作業の実装例と、その拡張戦略について述べる。本システムでは、安価なレーザーセンサおよび深度情報を取得可能なセンサを用いることで、産業用ロボットが苦手とする接触を伴う作業を回避しつつ、安定

した作業実行を可能としている。これにより、高価な専用設備に依存せず、実用的なロボット作業の実現を目指している。

対象物は工業生産物に限定し（図11）、個別対象物の認識結果から、対象物の形状および持点を生成AIが推定する構成を採用している。具体例として、ユーザーが「青い電線を把持して横に移動させる」といった自然言語指示を与えると、AIが対応する処理結果を生成する。生成結果がユーザーの意図と完全に一致しない場合には、対話的な修正を行なながらプログラムを生成する方式としている。

### 工業生産物

- ・形状種類多い

- ・1 mm程度の位置精度必要



図11 対象物

本方式により、タクトタイム約4秒の作業に対し、教示から動作開始までの時間は約20分程度で完了した。画像処理においては、多数のフィルタリング処理や複雑な前処理を行わず、生成されたプログラムを無修正のまま使用しても、ロボットが所定の作業を実行できることを確認している。さらに、眼鏡部品のような比較的微細な工業製品についても、所定の位置への配置作業が可能であることを実証しており、高精度なハンドリング性能を有することを確認している。

ただし、本システムが現時点で対象としている作業はピックアンドプレイスに限定されている。そのため、多様な作業を包括的に扱うことを目標とする大規模な基盤モデルと直接比較することは適切ではない。基盤モデルはより広範な作業への対応を志向しており、現段階で両者を同列に評価することはフェアではないと考えられる。

そこで、本研究では、個別作業単位で実現された知能を段階的に統合していく戦略を提案している。具体的には、ピックアンドプレイス作業のように、ユーザー入力からロボット動作までを一貫して実現可能な個別タスクをまず構築する。次に、挿入作業や組立作業など、異なる特性を有する別種の作業についても、それぞれ独立した形で実現を図る。その上で、これら複数の個別タスクを統合し、より大規模なシステムへと発展させることを目指す。

この統合段階においては、対象物認識および運動生成の統合機能が重要となる。これらを段階的に構築することで、巨大な単一基盤モデルとは異なるアプローチから、同等の機能性に対抗し得る戦略となる可能性がある。本研究では、この考え方を

「Individual Task Intelligence Integration」と位置づけているが、現時点では構想段階にあり、今後の研究開発によって実現可能性を検証していく必要がある。

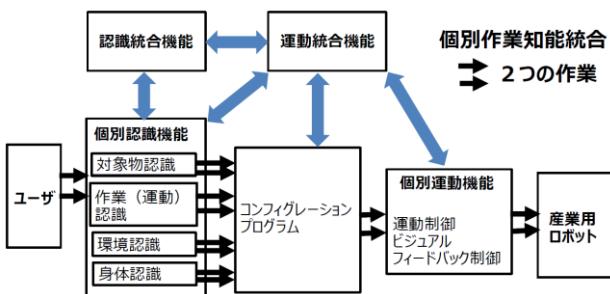


図 12 より複雑な状況または二つ目の作業実現のアプローチ

## 7.まとめ

本稿は、人口減少に伴う労働力不足という社会的背景の下、産業用ロボット分野における生成AI活用の現実的かつ短期的な解として、高精度なロボット動作自動生成手法を提案するものである。従来、産業用ロボットは教示再生方式により運用されてきたが、この方式は熟練SIerやティーチングマンへの依存度が高く、環境変動への適応性にも乏しいという課題を有している。一方、近年注目される基盤モデルによるエンドツーエンド制御は高い汎用性を期待できるものの、膨大な学習データを必要とし、産業用途で要求される精度・信頼性・実時間性を1年程度の短い研究開発期間で達成することは困難と予想される。

これらを踏まえ、著者らは生成AIをロボットシステムの中核ではなく「限定的な機能部品」として位置づけ、AI依存度を意図的に抑制した分割型アーキテクチャを採用する。提案手法の基盤技術として、絶対座標系の高精度同定を不要とし、ロボット先端と目標との相対誤差の収束のみで動作を実現する独自のビジュアルフィードバック制御を用いる。本手法は運動学や動力学モデルに依存せず、カメラキャリブレーション誤差や設置ズレを許容しつつ、約20μm級という産業用途に十分な高精度を達成できる点に特徴がある。

ロボティクスに内在する「知能・身体・環境」が相互に影響し合う三体問題に対し、本研究では明確な制約条件を導入することで問題を単純化する。身体は産業用ロボットアームと市販ハンドに限定し、

知能は汎用性を追求せず、まず需要の高いピックアンドプレイス作業に特化する。環境条件も水平面間移動や障害物なしといった単純な状況から扱い、探索的接触など産業用ロボットが苦手とする要素は排除する。

提案システムでは、ユーザーから自然言語および画像による入力を受け、機能を「対象物認識」「作業・運動認識」「環境認識」「身体認識」の四要素に分解する。このうち生成AIは対象物認識と作業・運動認識にのみ適用され、環境認識および身体認識には決定論的な従来技術を用いることで、再現性と安定性を確保する。生成AIは最終的にC++コード断片を出力し、非AIベースのプログラム統合機構を通して実行可能なロボット制御プログラムへと変換される。

実装例として、比較的低コストなハードウェア構成によるピックアンドプレイス作業を示し、教示から動作開始まで約20分、タクトタイム約4秒という実用的な性能を確認している。微細な工業製品に対しても高精度な配置が可能であり、生成されたプログラムを大幅な調整なしに使用できる点が示された。

本研究は、単一の巨大な基盤モデルを構築するのではなく、個別作業単位で実現した知能を段階的に統合していく「Individual Task Intelligence Integration」という戦略を提案する。これは、短期的な実装可能性と産業適用性を重視した現実的アプローチであり、生成AIとロボティクス技術を組み合わせた次世代産業ロボットシステムの有効な方向性を示すものである。



ロボット革命・産業 IoT イニシアティブ協議会  
Robot Revolution & Industrial IoT Initiative