

自律分散型画像センシングモジュール群からなる ロボットの知覚システムの検討

Perception System for Robot with Distributed Autonomous Image Sensing Modules

齋藤 俊太*¹ 青木 義満*¹
Shunta SAITO Yoshimitsu AOKI

*¹慶應義塾大学大学院理工学研究科

Graduate School of Integrated Design Engineering, Keio University

Robots in our living space or commercial facility require adaptive decision making system. To perform various tasks in those scenes flexibly, many essential recognition systems are necessary. We attempt to build a distributed autonomous image sensing modules to provide various information that can be extracted from environment in parallel. We consider how to apply the ontology-based intelligence system to real-world.

1. 研究概要

主に人とインタラクションすることを目的としたロボットにとって、非限定的な状況下での適応的な行動選択は重要な課題である。このためには周囲の瞬間的な状況だけでなく一連の事象をエピソードとして認識し、コンテキストを把握したり、人との対話が必要な場面では相手の特徴だけでなく表情や視線などの短時間しか表出しにくい細かな状態や行動を取得することができる認識機能が必要となる。また、ロボットが生活空間や商業施設などの人が行動するためにデザインされた空間で動作するためには、人だけではなく物体や空間にも注意を向ける必要がある。このような様々な対象は、その対象ごとに異なる時間スケールで状態を変化させている。さらに対象ごとに認識の困難さが異なるため、入力から認識結果を出力するまでに必要な処理時間も様々である。

多様な認識対象から得られる情報を元に行動決定を行うために、必要に応じて適宜どういった対象を認識すべきかを切り替えるという方法が考えられるが、何が必要かを知るための抽象的な認識機能の実現は難しい。そこでロボット本体または空間に設置されたセンサからの入力を多数の計算機に送信し、それぞれの計算機で個別対象ごとの認識処理を自律分散的に行っておき、必要な環境や人物の情報が事前に定義されたタスクフローや、行動シナリオ選択のためのルールに応じて、基本的には互いに独立に作動している自律分散型認識モジュールから必要な認識結果を随時取得できるようにしておく。

図1はこの様子を模式的に表したものである。まずロボットが導入される場所において導入を要求した者が与える主たる目的を「サービス」と定義する。例えば喫茶店業務を行うロボットであれば接客と配膳が主なサービスとなる。サービスはどのようなプロセスを実行することを繰り返すことで成り立っているかを事前に定義しておき、サービスの選択が行われると認識モジュールの結果や事前定義されたサービスの情報をもとに現在従うべきプロセスが自動生成されるような知能処理がオントロジ等を活用した Web サービスとして用意されているものとする。例えば客をテーブルに案内するというプロセスは、来店者の検出、人数の認識、性別や年齢などの属性認識、などの認識タスクに分けられ、その結果に応じて案内先として適している候補の中で空いているテーブルを探すといった別のタ

スクへ進んでいくという形でタスクフローとして生成される。これらの各タスクではどのような情報が必要かが予め分かっており、必要なモジュールへアクセスしてすでに同時並列的に行われていた計算の結果を取得するだけでフロー上を進んでいくことができる。

また、サービスごとにあり得る危険な状態や避けるべき状況があらかじめ知識として蓄えられているとすると、すべてのタスクと並列して常にそれを中断しなければならないかどうかを判断する状態チェックが行われていけばよい。状況のチェック機能においても前述のタスクの場合と同様に、判断を行うために必要な情報が何であるかが事前に定義されており、それに従って必要なモジュールへアクセスして計算結果を取得すればこのままタスクフローを進めて良いかがどうか決定可能なようにしておく。

このような構成においてはロボットの知覚は無意識的な自律分散型モジュールの集合として存在しており、これらを個別に分離して開発していけばよい。ただし各モジュール間で互いの結果を利用するような場合もあるため、特に高次の認識を行うモジュールは入力として用いる他のモジュールの計算コストがボトルネックとなり認識結果の更新頻度が制限されることが考えられる。

2. 構成モジュール

本研究では Aldebaran 社の NAO および Pepper を用いてサーバ・クライアント型のクラウドロボットシステムの構築を目指している。ロボットが外部の情報を得るためのセンサデバイスとしては、主に RGB カメラとデプスセンサから成る画像入力装置群を用いる。これまで画像を用いた物体や人物の認識技術についてはコンピュータビジョン分野で活発に研究が行われてきており、まずは既存の手法をモジュール化して並列に処理させることを目指す。現状では物体検出手法として R-CNN[Girshick 14] を用い、入力画像中で人物が検出された場合はその領域を入力として DeepPose[Toshev 14] を適用し姿勢推定を行う。また GoogLeNet[Szegedy 14] による物体認識手法もモジュール化しておき、姿勢推定モジュールの結果と合わせて所持物体認識モジュールを作成した。これは入力として手関節位置の周囲の画像を取り、これを物体認識モジュールの入力として来店した客が持っている荷物を認識するモジュールである。

連絡先: 齋藤 俊太, 慶應義塾大学, 神奈川県横浜市港北区日吉
3-14-1, 045-566-1796, ssaaito@aoki-medialab.org

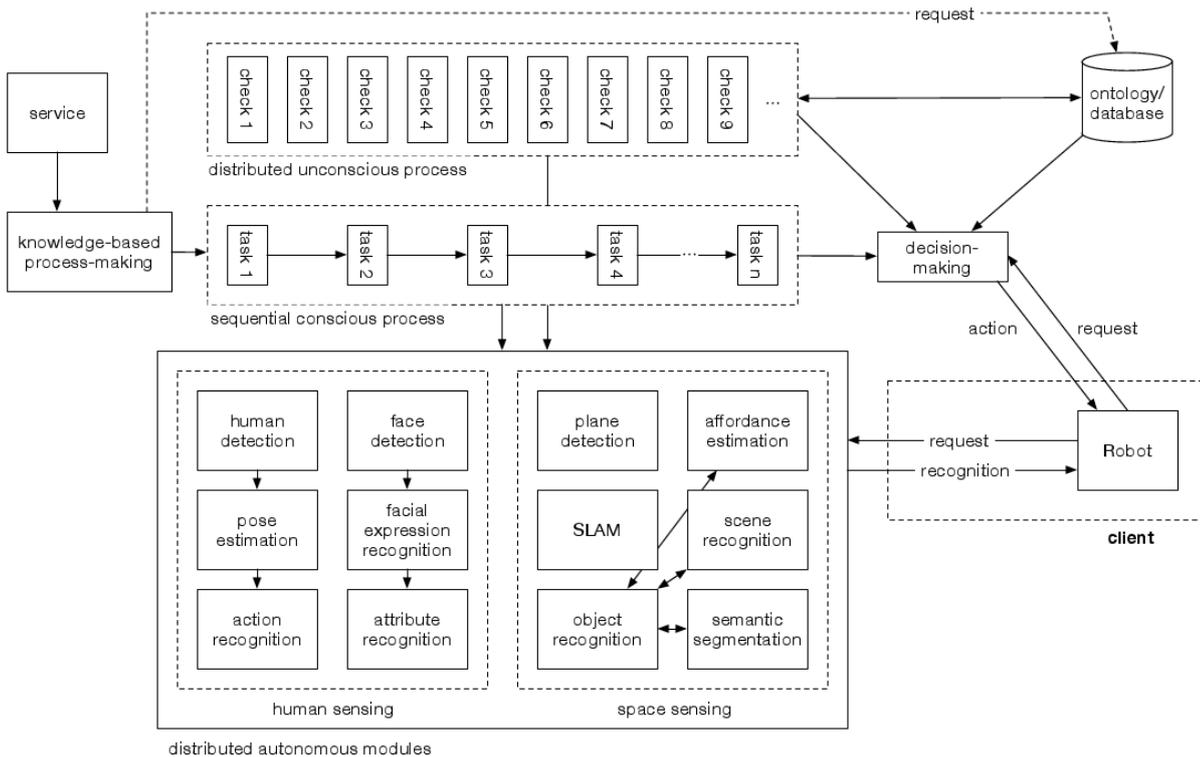


図 1: Distributed autonomous image sensing modules

3. 今後の課題

現状では既存の手法をただモジュールとして並列に実行可能にしたのみである。また、推論や行動決定を行う知能処理部分はオントロジの知識を応用したデータ構造となっているため、画像認識モジュールの出力結果をオントロジと対応付けるための中間処理が必要となっている。オントロジを構成する要素やプロパティに画像認識モジュールの出力を対応付けることを考える中で、オントロジとして表現された知識を元にモジュールの認識結果としてあり得る候補を制限したり、また現実空間からオントロジを更新するための情報取得を自動的に行なったりといった展開も考えたい。一方、物体認識は他クラス分類問題として扱われることが多いが、意味空間を大量のデータから獲得し意味ベクトルへの回帰問題として物体認識を捉える研究なども行われている [Frome 13]。Zero-shot learning 問題はロボットにおいて未知環境への適応を考える際に非常に重要であるし、物体認識を行った結果をどのように表現するか、という部分への考察を進めることでオントロジ分野で培われた知識の体系化方法を画像認識の問題設定の中に取り入れていくことも考えていきたい。

謝辞

本研究の一部は、科学技術振興機構 (JST) 戦略的想像研究推進事業 (CREST) 「実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践」の支援によって実施した。

参考文献

- [Girshick 14] Girshick, Ross and Donahue, Jeff and Darrell, Trevor and Malik, Jitendra: Rich feature hierarchies for accurate object detection and semantic segmentation, Computer Vision and Pattern Recognition (CVPR), 2014
- [Toshev 14] Toshev, Alexander and Szegegy, Christian: DeepPose: Human Pose Estimation via Deep Neural Networks, Computer Vision and Pattern Recognition (CVPR), 2014
- [Szegegy 14] Szegegy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew: Going Deeper with Convolutions, ILSVRC, 2014
- [Frome 13] Andrea Frome and Greg Corrado and Jon Shlens and Samy Bengio and Jeffrey Dean and Marc'Aurelio Ranzato and Tomas Mikolov: DeViSE: A Deep Visual-Semantic Embedding Model, Advances In Neural Information Processing Systems (NIPS), 2013