

# 多品種商品識別機能を有する倉庫内自動棚入れ棚出しロボットの開発

## – Amazon Robotics Challenge 2017参加報告 –

Development of Autonomous Stowing and Picking Warehouse Robot with Multiple Item Recognition Capabilities

- Participation Report on The Amazon Robotics Challenge 2017 -

岡田 征剛  
Seigo Okada

陽 品 駒  
Pin Chu Yang

岡崎 安直  
Yasunao Okazaki

浅井 勝彦  
Katsuhiko Asai

山本 正樹  
Masaki Yamamoto

グスタボガルシア  
Gustavo A. Garcia Ricardez

### 要 旨

Amazon Robotics Challenge (ARC) は近年の物流現場における急速な省人化・自動化の需要増加を受け、最重要課題である商品の識別および把持をテーマとして開催されるロボット国際競技大会である。ARC2017では、日々大量の新商品が追加される実際の物流現場を想定し、事前に発表される既知物体に加え、競技開始45分前に発表される未知物体をどう認識し、把持するかが重要なポイントとなった。筆者らは、本大会に奈良先端科学技術大学院大学と合同で出場した。物体認識手法として、ディープラーニングを用いた画像認識と特徴量をベースとした複数の識別手法を組み合わせ、各結果の確からしさを考慮して最終的な決定をするという手法を開発し、16チーム中6位の成績を取めた。

### Abstract

The Amazon Robotics Challenge (ARC) is an international robotics competition to accelerate the development of object recognition and picking automation technology in the logistics industry. At the ARC 2017, competitors were required to select specific items from a large number of items, with points awarded for correct identification and manipulation. To simulate real logistical environments where several new products are added daily, data on some of the items were provided to competitors before the challenge, and other items were unveiled 45 min before the start of each competition round. We used multiple recognition methods, including object recognition based on deep learning and feature matching, whose confidences were considered in determining the final recognition performance. We teamed up with the Nara Institute of Science and Technology for this competition, and ranked 6th among 16 teams.

## 1. はじめに

近年、物流現場において、eコマースの拡大や少子高齢化を背景とした労働力不足が懸念されており、物流現場の省人化・自動化ニーズが急速に高まっている[1]。

Amazon.com, Inc.の物流倉庫では、Kiva Systems (Amazon Robotics) の自律移動ロボットKiva Podが、商品が収納されている棚を自動搬送している。一方、棚入れおよび棚出し作業は人手によって行われており、自動化が期待されている。しかしながら、eコマースでは多品種の商品が棚にバラ積みで置かれているため、倉庫内の多種多様な商品を識別し、把持して所定の位置に移動させることが課題として挙げられる[2]。さらに、毎日大量の多品種の新商品が追加される状況においては、新しい商品にどのように対処するかも課題となる。Amazon.com, Inc. はこれらの課題解決を加速するため、2015年よりロボット国際競技大会であるAmazon Picking Challengeを開催してきた。

第3回目となる本大会はAmazon Robotics Challenge (ARC) という名称に変更され、2017年7月25日～30日に

日本の名古屋で開催された。筆者らは、本大会に奈良先端科学技術大学院大学 (NAIST) と共同で出場した。

本稿では、ARC 2017で要求された課題に対し、筆者ら NAIST-Panasonicチームがどのようなシステムを構築して臨んだか、特に多品種の商品を識別するにあたり、どのような認識戦略を取ったかについて述べる。

## 2. ARC2017の競技課題

ARC 2017では、日々大量の新商品が追加される実際の物流現場を想定し、大会数箇月前に発表される商品 (既知アイテム) に加え、競技開始45分前に初めて提示される商品 (未知アイテム) を競技で扱うこととなった。競技開始45分前に与えられるのは、対象アイテム、対象アイテムの重さおよび寸法データ、1つのアイテムにつき数枚の画像データである。

これらのデータおよび45分という短い時間内で収集したデータから、未知アイテムをどのように認識するかが本大会の大きな鍵となった。競技課題は、商品の棚入れ

に相当するStow Task, 棚出しに相当するPick Taskからなる。各タスクでは、正確性、早さに対してポイントが与えられ、このポイントが上位のチームが決勝ラウンドに進出する。決勝ラウンドではこれらのタスクを連続して行う。以下に各タスクの概要と評価方法を示す。

## 2.1 Stow Taskの概要と評価方法

Stow TaskではToteと呼ばれる箱に無造作に入れられた20個のアイテムをピックアップし、棚に収納する。この20個のアイテムは、10個の既知アイテムと、10個の未知アイテムとで構成される。全アイテムは重なり合った状態でTote内に入れられている。さらに、競技終了後、ピックアップしたアイテムがどの棚に収納されたかを自動的にファイルに出力して提出する必要がある。

アイテムを正しく棚に収納することができると、既知アイテム1個につき5ポイント、未知アイテム1個につき10ポイントが加算される。競技時間は15分間であり、すべてのアイテムを収納したうえで時間が残っていれば、5秒につき1ポイントが付与される。ファイルに出力したアイテムの場所が間違っていれば1アイテムにつき5ポイントが減算される。アイテムの大きな破損は20ポイントの減点、床への落下は15ポイントの減点となる。

## 2.2 Pick Taskの概要および評価方法

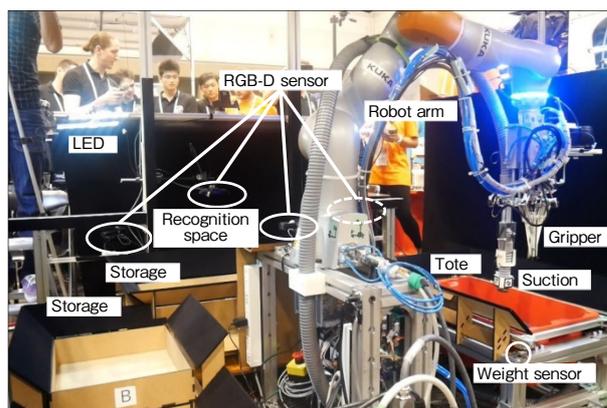
Pick Taskではあらかじめ棚に置かれた32個のアイテムのうち、要求された10個のアイテムを、配送用段ボールを想定した3つの箱のうち、指定された箱に収納するタスクである。アイテムを正しく箱に収納することができると、既知アイテム1個につき10ポイント、未知アイテム1個につき20ポイントが加算される。収納する箱を間違えると15ポイントの減点、要求されていないアイテムを箱に収納しても15ポイントの減点となる。

## 3. システム構成

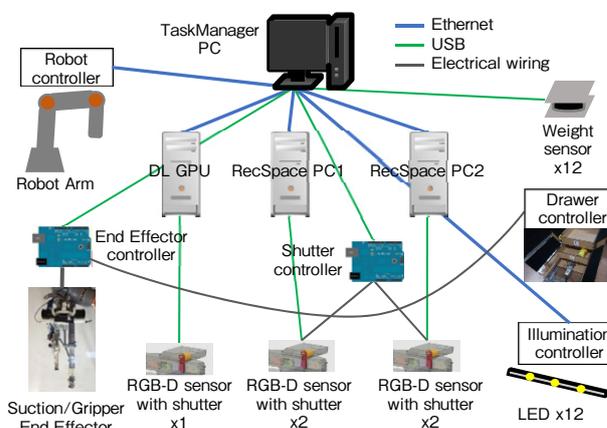
ソフトウェア開発では、ロボット研究の国際標準となりつつある、ROS<sup>(注1)</sup>を採用した。

ロボットシステムの全体図を第1図、システムの構成を第2図に示す。産業用7軸多関節アームの先端に空気圧式の把持機構および吸着機構からなるエンドエフェクタを搭載した。アームは14 kgまで把持が可能であり、吸着に用いる真空圧は産業用ブLOWERにより、-40 kPaまで出力可能である。

センサ系は、アーム先端部にRGB-Dセンサを配置し、



第1図 ARC2017で用いたシステム  
Fig. 1 Proposed solution at the ARC 2017



第2図 システム構成  
Fig. 2 Deployment diagram

「認識空間」と呼ぶ棚上部 (Recognition space, 第1図) 周辺に4台のRGB-Dセンサを配置する。RGB-DセンサはRGBセンサとDepthセンサが一体となっており、ROSドライバが公開されているものを選定した。Depthセンサはプロジェクタからパターンを投影して距離を算出する方式で、IRレーザ光をアクティブに発するため、相互干渉が生じないように小型のメカシャッターを搭載した。

棚はアイテムの重なり合いを最小限にして物体認識を容易にする目的で面積を広く設計するとともに、総面積を規定するルールに準ずる形で、小物部品を専用に収納するための自動引き出しも設置した。さらに、3つの棚の下には各4個の重量センサを配置し、物体認識に利用した。さらに、照明環境を一定にするため、外光をなるべく遮断したうえで、LED照明により明るさを制御した。

## 4. アイテム識別戦略

本章では、未知アイテムに対応するための戦略、およ

(注1) Robot Operating System (ROS). <http://www.ros.org/>

びアイテムを認識するにあたってどのような手法を用いたかについて詳細に説明する。

#### 4.1 未知アイテムへの対応

第3図に競技会前、競技直前および競技中の認識に関わる処理の流れを示す。既知アイテムに対しては、競技会前に学習を完了している。未知アイテムに対しては45分間でデータ収集および学習を完了させる必要があった。そのため未知アイテムに対してディープラーニングを行うには時間が足りないと考え、他の認識手法を用いた。

本大会のタスクでは、Toteの中にアイテムが積み重なった状態でアイテムを認識する必要があり、画像中に対象物の特徴が入り混じった状態のため、物体認識には不利な条件であった。ただし、4.2節で述べるディープラーニングによる画像認識は上記の条件においてもアイテムを認識することが可能であり、さらに未知アイテムに対してもアイテムの場所を検出することが可能であった。そのため、競技中はまずディープラーニングによって検出したアイテムをピックアップして認識空間へ移動させ、認識が容易な条件であらためて他の認識をするという手段を取った。

これら複数の認識システムを並行して用いたうえ、各認識システムから得られる、認識の信頼度を候補となる全アイテムに対して算出し、各信頼度から重み付け投票

により最終的な認識結果を算出するという手法を取った（重み付け投票システム、第4図）。複数の認識システムを用いることで全体としてロバストな認識システムとなることが期待されたため採用した。次項以降に各詳細を述べる。

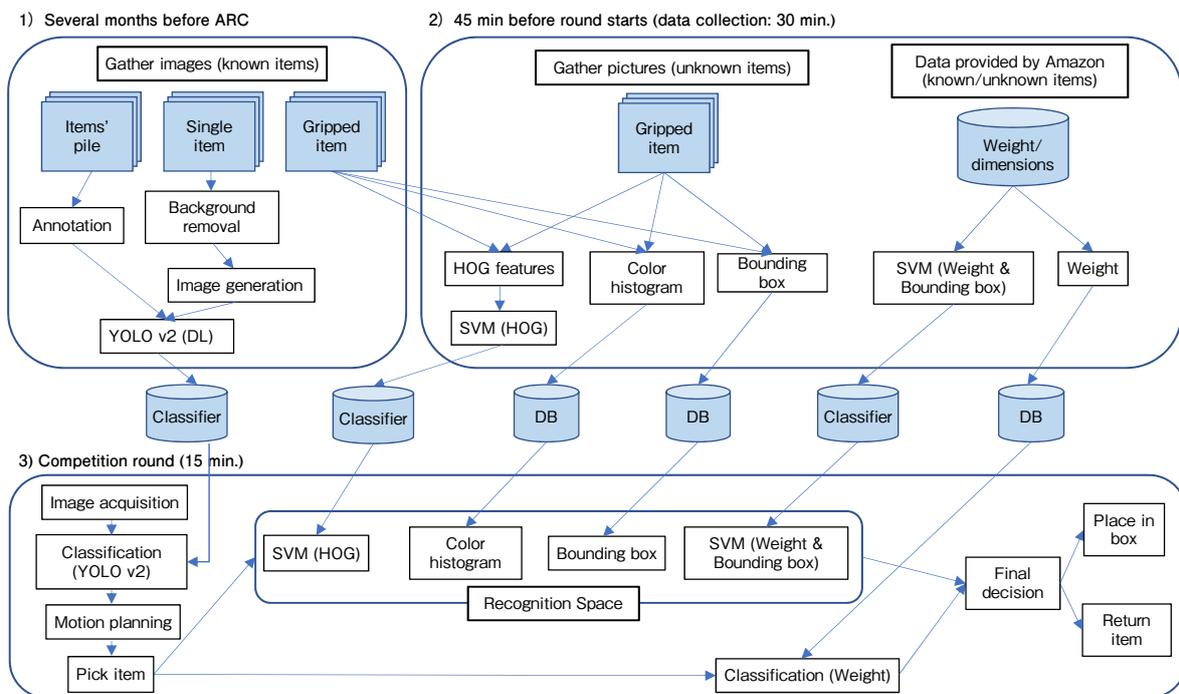
#### 4.2 ディープラーニングによる画像認識

ディープラーニングによる画像認識（第4図、(a)）は、多クラスの物体を高速で識別可能なYOLO v2[3]を用いた。フレームワークはTensorFlow<sup>(注2)</sup> [4]を用いた。

学習の際には、まず既知の40アイテムに対して、アイテム1個ずつを撮影し、各画像から背景を除去した。次に、棚およびToteの何も入っていない状態の画像の上にこれらの背景を除去した画像を合成していき、実際の競技における状況に近い画像を生成し、これらを学習データとした。さらに、棚およびToteの中で数のアイテムが重なり合った状態を撮影し、これらの実画像データに追加学習させた。

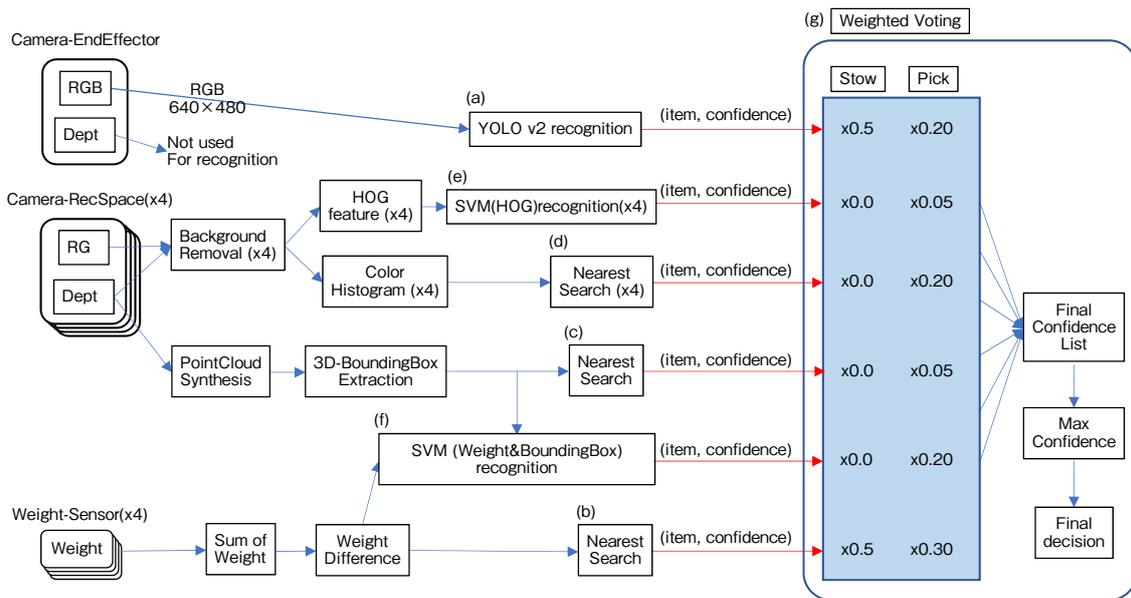
全体で約8000枚の実画像データを収集し、人手によるアノテーションを行った。さらに、それらにノイズや色の変化を加えることで約50000枚の画像データを生成し、学習データとした。

(注2) Google Inc.の登録商標または商標。



第3図 各時間軸における処理の流れ

Fig. 3 Task flow at each time frame



第4図 認識システム

Fig. 4 Recognition system

### 4.3 重量センシングによる認識

重量センシングによる物体識別（第4図, (b)）は、棚を下から4点で支える形で重量センサを配置し、棚からアイテムが取り除かれた際の重量の変化と、競技開始45分前に与えられる重量データを比較して、取り除かれたアイテムを推定するという手法を用いた。重量変化検出の精度は10 gであった。

### 4.4 認識空間における認識

棚内のアイテムをディープラーニングによる画像認識（第4図, (a)）でピックアップすると、ロボットアームは棚上部の「認識空間」へアイテムを移動させる。ここでは、認識空間で用いられるセンシング手法（第4図, (c)～(f)）について述べる。

#### [1] バウンディングボックスマッチング

認識空間では4方向から距離画像を撮影して3次元の点群を生成し、主成分分析により直方体をフィッティングすることで、アイテムの3次元バウンディングボックスを生成し、バウンディングボックスの縦横高を特徴量として用いてマッチングすることによって物体を識別した。マッチングは特徴量を表すベクトルのユークリッド距離を算出することによって行った（第4図, (c)）。

#### [2] カラーヒストグラムマッチング

認識空間で得られた4方向からのRGB画像から、距離情報を用いてアイテム以外の背景を黒にする処理を行ったうえで、カラーヒストグラムを生成し、それを特徴量として、バウンディングボックスマッチングと同様の方法

でマッチングすることによって物体を識別した（第4図, (d)）。

#### [3] HOG-SVMによる識別

SVM (Support Vector Machine) [5]は機械学習の1種であり、ディープラーニングのように学習時間がかからないため、45分以内でデータの取得および学習が可能であった。

認識空間で得られたRGB画像から距離情報を用いて背景を削除した画像の特徴量を算出し、SVMを用いることで識別を行った（第4図, (e)）。画像の特徴量には、SIFT[6]やORB[7]などさまざまな画像特徴量を用いてテストした結果、最も精度が良かったHOG (Histogram of Oriented Gradients) [8]特徴量を用いた。なお、HOG特徴量は回転にロバストではないため、主成分分析によりアイテムの向きを揃（そろ）える処理を行った。

#### [4] 重量・バウンディングボックスSVMによる識別

4.3節および[1]項で得られたアイテムの重量情報およびバウンディングボックス情報を特徴量としてSVMに用いることで識別を行った（第4図, (f)）。SVMの学習には競技開始45分前に与えられる重量および寸法情報を用いた。

### 4.5 重み付け投票システム

本節では「重み付け投票システム」について述べる。4.2節から4.4節で示した各認識システムはそれぞれ精度が異なるため、精度が高いシステムほど重みを増やし最終結果に反映させた（第4図, (g)）。

Stow Taskではどのアイテムをどの棚に収納したかというファイルを自動的に生成する必要があるが、アイテムの誤認識による減点よりも移動させたアイテムの数による加点を重視し、認識精度よりも速度を優先するという戦略を取った。そのため、1個のアイテムにつき約5秒要していた認識空間における認識は使用せず、ディープラーニングおよび重量センシングのみを認識に用いた。

Pick TaskではStow Taskとは異なり、すべてのアイテムを正確に認識する必要があると考えた。ディープラーニングによる認識は既知の物体しか学習していないため、投票の重みはStow Taskよりも低く設定し、4.2節から4.4節で示した複数の認識システムで認識を行った。

## 5. 結果と考察

予選ラウンドのStow Taskで110ポイント、Pick Taskで80ポイントを取得して決勝ラウンドに進み、決勝ラウンドではStowTaskで40ポイント、PickTaskで50ポイントの、合計90ポイントを取得して6位という成績を収めた。

アイテムの認識に注目すると、以下の利点および欠点があった。

- 1) ディープラーニングによる画像認識
  - ・ 既知アイテムに対する高い認識精度
  - ・ 未知アイテムに対してもアイテムの場所を検出可能
  - ・ 45分以内でのデータ取得および学習は困難
- 2) 重量センシング
  - ・ 重量変化検出精度 (10 g) の範囲内に候補となるアイテムが複数存在しない場合は正しく認識可能
  - ・ 重量データは運営側から提供されるためデータ取得不要
- 3) バウンディングボックスマッチング
  - ・ アイテムが変形しやすい場合、学習したバウンディングボックスと形状が大きく異なってしまう誤認識
- 4) カラーヒストグラムマッチング
  - ・ 特徴的な色のアイテムに対しては認識精度が高い
- 5) HOG-SVMによるアイテム識別
  - ・ 候補アイテムに対する認識結果の信頼度を降順に並べた際、対象となるアイテムの信頼度が上位になる確率が高いが、最上位になる確率は低い
- 6) 重量・バウンディングボックスSVM
  - ・ 重量と縦横高というアイテムの2種類の情報を利用するため認識精度が高い
  - ・ 学習に用いるデータは運営側から提供されるためデータ取得不要

総じて、アイテムの色情報や、重量情報およびバウンディングボックス情報を組み合わせた情報が本選での成

績に貢献したと考えられる。Pickタスクにおいて、予選ラウンドで8個、決勝ラウンドで11個のアイテムを認識し、未知アイテムを含めてすべて正しく認識されたことから、投票システムによりロバストな認識を実現できたと考えられる。

一方、優勝チームであるACRV (Australian Centre for Robotic Vision) は独自で開発したネットワーク構造により、未知アイテムに対してもディープラーニングを適用するだけでなく、物体の形状をセグメンテーションする手法により、精度の高い認識や把持を実現していた。[9]

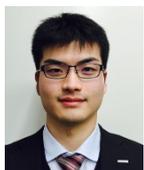
## 6. まとめ

本稿で述べた手法を用いて、初出場ながら6位という成績を収めたことは大きな成果である。一方、今回用いた手法では、Pick Taskにおいて認識空間で認識を行うまで認識結果を出せないことから、時間のロスが生じた。今後、ACRVのように認識空間を用いずに認識する手法の導入を検討するとともに、信頼性やコストの観点含めどのように実用化に近づけていくかが課題である。

## 参考文献

- [1] 森山 和道, “物流ロボットまとめ, 「圧倒的な人手不足」による逆境克服から学ぶべきこと,” ビジネス+IT, <https://www.sbbit.jp/article/cont1/32852>, 参照 Apr. 20, 2018.
- [2] 藤吉弘亘 他, “[特別講演] Amazon Picking Challenge 2016の参加レポート,” パターン認識・メディア理解研究会, vol.116, no.461, pp.123-129, 2017.
- [3] J. Redmon et al., “Yolo9000: Better, faster, stronger,” eprint arXiv:1612.08242, Dec. 2016.
- [4] A. Martin et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” eprint arXiv:1603.04467, Mar. 2016.
- [5] C. Cortes et al., “Support-Vector Networks,” Machine Learning, vol.20, issue 3, pp.273-297, Sept. 1995.
- [6] D.G.Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol.60, issue 2, pp.91-110, Nov. 2004.
- [7] E. Rublee et al., “ORB: An efficient alternative to SIFT or SURF,” 2011 International Conference on Computer Vision, Barcelona, pp. 2564-2571, Nov. 2011.
- [8] Dalal. N et al., “Histograms of Oriented Gradients for Human Detection,” IEEE CVPR, pp.886-893, June 2005.
- [9] 進藤智則 他, “Amazon Picking Challenge 2017 詳細報告【前編】,” 日経ロボティクス, no.27, pp.5-17, Oct. 2017.

## 執筆者紹介



岡田 征剛 Seigo Okada  
ビジネスイノベーション本部  
AIソリューションセンター  
AI Solution Center, Business Innovation Div.



陽 品駒 Pin Chu Yang  
ビジネスイノベーション本部  
AIソリューションセンター  
AI Solution Center, Business Innovation Div.



岡崎 安直 Yasunao Okazaki  
ビジネスイノベーション本部  
AIソリューションセンター  
AI Solution Center, Business Innovation Div.



浅井 勝彦 Katsuhiko Asai  
ビジネスイノベーション本部  
AIソリューションセンター  
AI Solution Center, Business Innovation Div.  
博士（工学）



山本 正樹 Masaki Yamamoto  
ビジネスイノベーション本部  
AIソリューションセンター  
AI Solution Center, Business Innovation Div.  
工学博士



グスタボ ガルシア Gustavo A. Garcia Ricardez  
奈良先端科学技術大学院大学 情報科学領域  
Division of Information Science,  
Nara Institute of Science and Technology  
Ph.D. in Robotics