

## [特別講演] Amazon Robotics Challenge 2017 の参加レポート

藤吉 弘亘<sup>†</sup> 岡田 慧<sup>††</sup> 江原 浩二<sup>†††</sup> Gustavo Garcia<sup>††††</sup>

<sup>†</sup> 中部大学 〒487-8501 愛知県春日井市松本町 1200

<sup>††</sup> 東京大学 〒113-8656 東京都文京区本郷 7-3-1

<sup>†††</sup> 東芝インフラシステムズ株式会社 〒212-8585 神奈川県川崎市幸区堀川町 72 番地 34

<sup>††††</sup> 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916 番地の 5

E-mail: <sup>†</sup>hf@cs.chubu.ac.jp, <sup>††</sup>k-okada@jsk.t.u-tokyo.ac.jp, <sup>†††</sup>koji.ehara@toshiba.co.jp,

<sup>††††</sup>garcia-g@is.naist.jp

**あらまし** e コマースにおける物流倉庫の自動化の課題は、棚に陳列された多品種の商品を識別し、把持計画により安定したピック&プレースを実現することである。Amazon.com, Inc. が 2015 年に開催した国際ロボット競技大会である Amazon Picking Challenge (APC) 2015 は、棚から商品を取り出して運搬用ボックスへ移動させる pick タスクが取り上げられた。そして、翌年に開催された APC 2016 では、pick タスクの他に運搬用ボックスから商品を取り出して棚へ収納する stow タスクが追加された。2017 年には、大会名が Amazon Robotics Challenge (ARC) に改名された。ARC 2017 は、競技直前に提供される新しい商品の認識・取り出しといったチャレンジングな問題設定が追加され、より実問題を想定した競技大会となった。本稿では、Amazon Robotics Challenge 2017 に参加した各チームの取り組みを紹介し、今後の技術課題について述べる。

**キーワード** Amazon Robotics Challenge, ピッキング, ロボット競技, 画像認識, 深層学習

### 1. はじめに

Amazon.com, Inc. の物流倉庫では、Kiva Systems (2018 年現在 Amazon Robotics) の自律移動ロボット kiva pod が、商品を収納した棚をピッキング担当者まで自動搬送する [1]。kiva pod の導入により、人間が倉庫内を移動して商品を探す必要はなくなるが、棚からの商品のピッキングは人間の作業である。そのため、ピッキングロボットによる物流倉庫の自動化が期待されている。e コマースでは、多品種の商品が棚に無造作に収納されているため、棚の中の商品を正確に識別し、把持計画により安定したピック&プレースを実現することが、物流倉庫の自動化に向けた重要課題である。このような背景から、Amazon.com, Inc. は物流倉庫の自動化を競う初の国際ロボット競技大会「Amazon Picking Challenge」を開催した。

2015 年 5 月に米国・シアトルにて開催された Amazon Picking Challenge (APC) 2015 では、多品種の商品のピッキング問題が取り上げられ、棚の bin と呼ばれる 12 箇所の枠の中から指定された 25 種類の商品 (アイテム) を取り出し、tote と呼ばれる運搬用ボックスへ移動させる pick タスクが競技対象であった。ピッキング対象アイテムの種類は、剛体、非剛体、半透明物体で構成されており、単純な画像認識アルゴリズムで解くことはできない問題設定となっていた [2]。

2016 年 7 月にドイツ・ライプツィヒにて開催された APC 2016 では、pick タスクに加えて tote から商品を取り出して棚へ収納する stow タスクが追加された。また、認識対象アイテム

も 25 種類から 39 種類へと増加し、物体把持と画像認識の 2 つの観点において難易度が高くなった。

2017 年 7 月に日本・名古屋にて開催された大会では、大会名が Amazon Robotics Challenge (ARC) に改名され、より実問題を想定した競技大会となった。まず、ARC 2017 では事前に提供される 40 種類のアイテムに加えて競技開始直前に提供される新しいアイテムを認識して取り出す課題が要求された。物流倉庫では、日々新しい商品が追加される。もし、機械学習ベースの画像認識システムを構築する場合、取り扱う商品が更新されるたびに画像認識システムを再学習するのは非現実的である。この問題を、どのように解決して効率的な画像認識システムを構築するかが ARC 2017 において重要課題である。さらに、pick タスクと stow タスクのスコアの合計が上位のチームは最終ラウンドとして stow タスクから pick タスクまでの一連の動作を競うタスクが追加された。また、ARC 2017 ではこれまで大会運営から提供されていた棚が廃止となり、storage と呼ばれる収納ボックスが導入された。Storage は各チーム独自に開発するルールとなっており、各ロボットシステムに適した storage の作成が必要である。

本稿では、Amazon Robotics Challenge 2017 に参加した各チームの取り組みを紹介し、今後の技術課題について述べる。

### 2. ARC 2017 の競技タスク

ARC 2017 で扱うアイテムは、事前に提供される 40 種類のアイテムと競技開始直前に提供されるアイテムがある。当日提

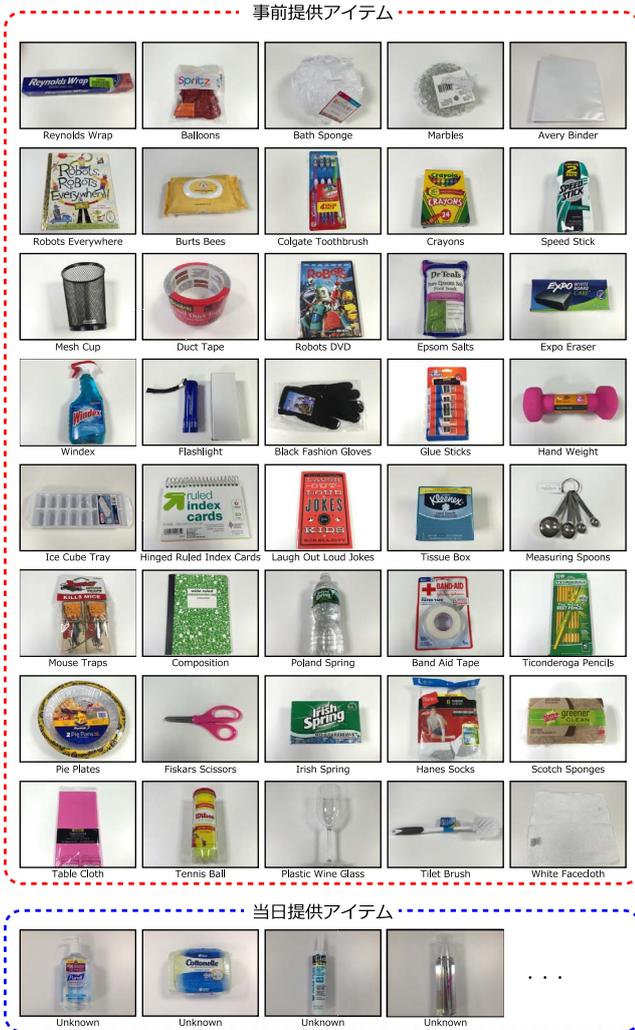


図 1 ARC 2017 での対象アイテム。

供アイテムは各競技タスク・各チームで全て異なる未知のアイテムが割り当てられる。当日提供アイテムは競技開始の 45 分前に受け取り、競技開始の 15 分前までの 30 分間は当日提供アイテムの画像撮影、重量計測、画像特徴抽出などの処理が許されている。競技本番では、これらの情報をもとに当日提供アイテムを認識しなければならない。図 1 に ARC 2017 の 40 種類の事前提供アイテムと当日提供アイテムの例を示す。以下に各タスクの概要と評価方法について述べる。

### 2.1 Stow タスク

Stow タスクでは、tote にばら積みされた 20 個のアイテムを storage へ収納するタスクである。Tote 内のアイテムは事前提供アイテム 10 個、当日提供アイテム 10 個の割合で配置される。Stow タスクにおける tote 内のアイテム配置例を図 2 に示す。Tote にばら積みされたアイテムをロボットが把持し、storage へ収納することで得点が得られる。Stow タスクでは、表 1 に従ってスコアが計算される。Stow タスクの競技時間は 15 分である。また、認識の信頼度が低いアイテムや識別不可と判断したアイテムなどは amnesty tote と呼ばれるボックスに収納することが許されている。Amnesty tote 内のアイテムはペナルティの対象外として扱われる。



図 2 Stow タスクにおける tote 内のアイテム配置例。

表 1 Stow タスクのスコア。

状況	得点
事前提供アイテムを storage に収納	5
当日提供アイテムを storage に収納	10
タスク完了時間のボーナス	残り時間 / 5 [s] (最大 180)
アイテムが外に残る	-15
認識アイテムの収納位置の誤り	-5
15 cm 以上の高さからアイテム落下	-5
アイテムが storage から 2 cm 以上はみ出す	-5
アイテムの傷・へこみ	-5
アイテムの大きな裂け目・穴・破砕	-20

表 2 Pick タスクのスコア。

状況	得点
事前提供アイテムを段ボール箱に収納	10
当日提供アイテムを段ボール箱に収納	20
タスク完了時間のボーナス	残り時間 / 5 [s] (最大 180)
タスク完了・段ボールが閉じられる	10
アイテムが外に残る	-15
認識アイテムの収納位置の誤り	-5
15 cm 以上の高さからアイテム落下	-5
アイテムが段ボール箱から 2 cm 以上はみ出す	-5
アイテムの傷・へこみ	-5
アイテムの大きな裂け目・穴・破砕	-20
段ボール箱に収納できないアイテムが存在	-25

### 2.2 Pick タスク

Pick タスクでは、storage に配置された 32 個のアイテムを発送用段ボール箱へピックアップするタスクである。Storage 内のアイテムは各チームで自由に配置できるが、競技開始直前に運営スタッフによりアイテムの位置・姿勢を変えられる。Storage 内のアイテムは事前提供アイテム 16 個、当日提供アイテム 16 個の割合で配置される。Storage に配置されたアイテムをロボットが把持し、発送用段ボール箱へピックアップすることで得点が得られる。Pick タスクでは、表 2 に従ってスコアが計算される。Pick タスクの競技時間は 15 分である。Stow タスクでは amnesty tote の使用が許可されていたが、pick タスクでは amnesty tote の使用が不可である。

### 2.3 最終ラウンド

ARC 2017 では、出場する 16 チームのうち、pick タスクと stow タスクのスコアの合計が 100 点以上かつ上位 8 チームが最終ラウンドにて競技を行う。最終ラウンドでは、stow タスクから pick タスクまでの一連の動作を競う。前半は stow タスクであり、16 個のアイテム(事前提供 8 個、当日提供 8 個)を競

チーム名	所属機関 (国)	ロボット	アームタイプ	エンドエフェクタ	ビジョンセンサ
ACRV	クイーンズランド工科大学 (豪) アデレード大学 (豪) オーストラリア国立大学 (豪) モナシユ大学 (豪)	自作 × 1	クレーン型	吸着, グリッパ	Intel 社 SR300
Applied Robotics	スマートロボティクス (蘭) シドニー大学 (豪)	Universal Robotics 社 UR5 × 1	単腕	吸着	Intel 社 SR300
GMU-Negev	ジョージメイソン大学 (米) ネゲヴ・ベン＝グリオン大学 (以)	Rethink Robotics 社 Baxter × 1	双腕	グリッパ	R200
IFL PiRo	カールスルーエ工科大学 (独) ヘルムートシュミット大学 (独)	Universal Robotics 社 UR3 × 1	単腕	吸着, グリッパ	DS325
IITK-TCS	インド工科大学カーンプル校 (印) タタ・コンサルタンシー・サービス (印)	Universal Robotics 社 UR5 × 1	単腕	吸着	MicroTechnica 社 ENSENSO
MIT-Princeton	マサチューセッツ工科大学 (米) プリンストン大学 (米)	ABB 社 IRB 1600ID × 1	単腕	吸着, グリッパ	Intel 社 SR300
NAIST-Panasonic	奈良先端科学技術大学院大学 (日) Panasonic (日)	KUKA 社 LBR iiwa × 1	単腕	吸着, グリッパ	Intel 社 SR300
Nanyang	南洋理工科大学 (新)	Universal Robotics 社 UR5 × 2	単腕	吸着	ZED Stereo Camera
NimbRo Picking	ボン大学 (独)	Universal Robotics 社 UR3 × 2	単腕	吸着, グリッパ	Photoneo 社 PhoXi 3D スキャナ XL
PLAID	カーネギーメロン大学 (米)	Universal Robotics 社 UR5 × 1	単腕	吸着	Photon
Team Duke	デューク大学 (米)	STAUBLI 社 ロボットアーム × 1	単腕	吸着	Intel 社 SR300
Team K	東京大学 (日)	Rethink Robotics 社 Baxter × 1	双腕	吸着, グリッパ	Intel 社 Orbbec Astra min
Team MC <sup>2</sup>	三菱電機 (日) 中部大学 (日) 中京大学 (日)	三菱電機社 MELFA RV-13FL × 2	単腕	吸着, グリッパ	三菱電機社 MELFA-3D Vision
Team T2	東芝 (日) 鳥取大学 (日)	川崎重工社 RS020N × 1	単腕	吸着, グリッパ	Intel 社 SR300
TKU M-Bot	淡江大学 (台)	自作 × 1	双腕	吸着, グリッパ	Intel 社 SR300
UJI RobInLab	ジャウメ1世大学 (西)	Rethink Robotics 社 Baxter × 1	双腕	吸着, グリッパ	Microsoft 社 Kinect v2

図 3 ARC 2017 の出場チーム一覧 (アルファベット順).

技直前に storage に配置させる。Tote 内にばら積みされた 16 個のアイテム (事前提供 8 個, 当日提供 8 個) をロボットにより storage へ収納する。スコア計算は表 1 と同じであり, 前半は amnesty tote の使用が許可されている。後半は pick タスクであり, 前半の stow タスクでアイテムを収納した状態の storage から, ロボットでアイテムを把持して発送用段ボール箱へ入れる。スコア計算は表 2 と同じであり, 後半は amnesty tote の使用が許可されていない。最終ラウンドの競技時間は, 前後半合わせて 30 分であり, 前半と後半の時間配分は各チームで自由に設定することができる。

### 3. ARC 2017 の競技結果

ARC 2017 では, 事前審査を通過した 16 チームが大会に出場した。事前審査では, 各チームの技術力を示すデモンストラーション動画とエントリーシートに基づいて出場チームが選ばれる。米国から 4 チーム, 日本から 4 チームの参加であった。今大会は日本での開催ということもあり, 日本国内からの ARC への注目を集めた。他の参加国は, イスラエル (以), インド (印), オーストラリア (豪), オランダ (蘭), シンガポール (新), スペイン (西), 台湾, ドイツ (独) であった。図 3 に全参加チームの詳細を示す。前大会の APC 2016 では, 双腕ロボットである Baxter が 6 台と多くのチームが使用していた。ARC 2017 では, Universal Robotics 社製の単腕ロボットアームが 6

台と多くのチームが使用した。エンドエフェクタには, 吸着タイプとグリッパタイプの 2 種類を装備し, アイテムによって使い分けるロボットが多く見られた。

#### 3.1 Stow タスクの競技結果

大会初日は stow タスクが開催され, 1 位は MIT-Princeton (160 点), 2 位は Nanyang (125 点), 3 位は Team MC<sup>2</sup> (120 点) であった。1 位の MIT-Princeton は, 図 4 に示すように storage と tote の両側に固定された 16 台のビジョンセンサで画像と距離データを取得する。16 台のビジョンセンサのうち 8 台のビジョンセンサで storage または tote 内の物体を撮影し, 物体の把持位置を検出する。Storage または tote 内の物体を把持し, ある程度の高さまで物体を持ち上げたときに, 残りの 8 台のビジョンセンサで把持物体を撮影して物体を認識する [3]。MIT-Princeton はロボットシステムや認識システムの完成度が非常に高く, stow タスクにおいて今大会で唯一全てのアイテムを把持して収納することに成功した。

#### 3.2 Pick タスクの競技結果

大会 2 日目は pick タスクが開催され, 1 位は Nanyang (257 点), 2 位は NimbRo Picking (245 点), 3 位は IITK-TCS (160 点) であった。1 位の Nanyang は, 図 5 に示すように産業用ロボットアーム 2 台を同時に動作させ, 効率的なピッキングシステムを実現した。ビジョンセンサは, 2 つの bin に仕切られた storage の真上に 1 台ずつ設置されており, エンドエフェクタ

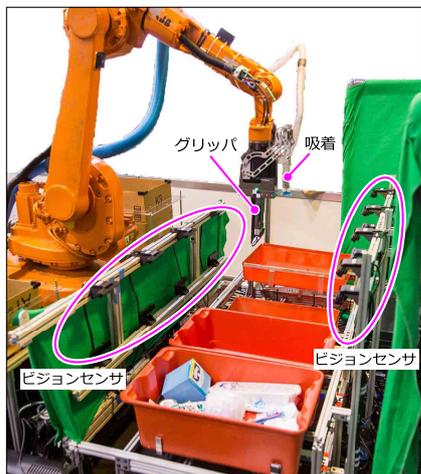


図 4 MIT-Princeton のロボットシステム。

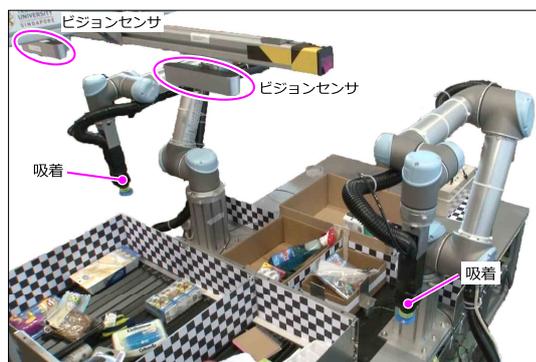


図 5 Nanyang のロボットシステム。

は吸着のみである。Storage は競技開始時にボックスの広さを拡張させる工夫が施されている。Storage を広くすることでアイテム同士の重なりが少なくなるため、ピッキング動作が非常に安定していた。

### 3.3 最終ラウンドの競技結果

大会最終日は最終ラウンドが開催された。最終ラウンドは stow タスクと pick タスクの合計スコアの上位チームが出場することができる。最終ラウンドの出場チームは、Nanyang, MIT-Princeton, IITK-TCS, NimbRo Picking, ACRV, NAIST-Panasonic, IFL PiRo, Applied Robotics の 8 チームである。最終ラウンドでは stow タスクから pick タスクへの一連の競技が行われ、1 位は ACRV (272 点)、2 位は NimbRo Picking (235 点)、3 位は Nanyang (225 点) であった。1 位の ACRV は、図 6 に示すようにクレーン型の自作ロボットを使用した。ビジョンセンサは、ハンドに固定されている。クレーン型のハンドには、グリッパと吸着パッドが水平に取り付けられており、把持アプローチの際にはハンドを 90 度または-90 度回転させる。Storage はボックスを 2 個並べたシンプルな構造であったが、物体把持の精度が高く、最終ラウンドでは高い得点を獲得した。

## 4. チーム紹介：Team MC<sup>2</sup>

Team MC<sup>2</sup> は、三菱電機グループ(三菱電機株式会社、三菱

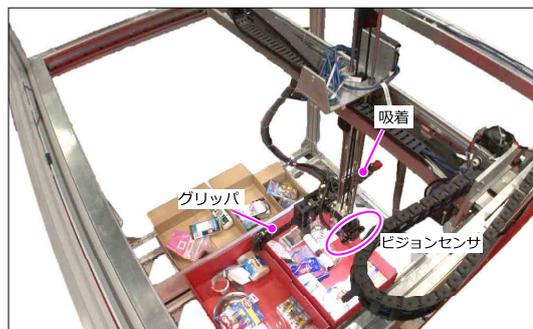


図 6 ACRV のロボットシステム。

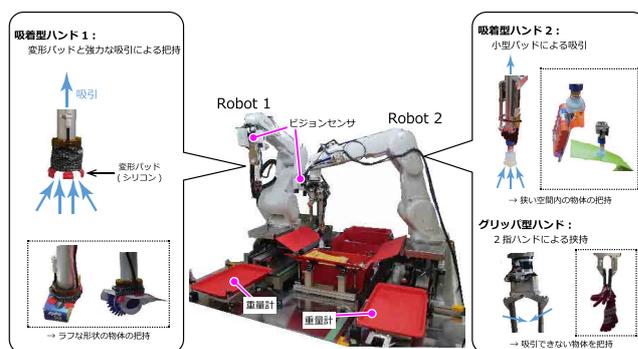


図 7 Team MC<sup>2</sup> のロボットシステム。

電機システムサービス株式会社、ターゲット・エンジニアリング株式会社)、中部大学 機械知覚&ロボティクスグループ(藤吉・山下研究室)、中京大学 知的センシング研究室(橋本研究室)により構成される合同チームである。本チームは 2015 年の第 1 回大会から連続して出場しており、今大会で 3 度目の出場となる。APC 2015 と APC 2016 における本チームの取り組みは文献 [4]~[6] を参考にさせていただきたい。以下に、Team MC<sup>2</sup> のシステム構成、ビジョン戦略、並びに特徴について述べる。

### 4.1 システム構成

Team MC<sup>2</sup> のロボットシステムは、図 7 に示すように 2 台の独立した産業用ロボットアームで構成される。ロボットアームは、それぞれ 13kg 可搬 MELFA RV-13FL を使用した。各ロボットのハンドにはビジョンセンサ (MELFA-3D Vision) と力覚センサを搭載している。力覚センサは、storage や tote などへの衝突回避に利用している。アイテムの重さを計測するために、重量計を搭載した仮置き台が設置されている。2 台のロボットは、1 軸の走行台に設置されており、storage ↔ tote/段ボール箱 ↔ 仮置き台 の間は走行台でロボットアームが移動する。

ロボットハンドは、吸着型とグリッパ型を使用した。Robot 1 には大型パッドの吸着型ハンドのみを搭載し(図 7 左)、Robot 2 は小型パッドの吸着型ハンドとグリッパ型ハンドの 2 種類を搭載した(図 7 右)。大型吸着ハンドは、吸引力が強く物体との接触面積が広いいためラフな形状で大きな物体の把持に適している。小型吸着ハンドは、パッドが小さいため、狭い空間内の物体や小さな物体の把持に適している。基本的に、吸着型ハンド

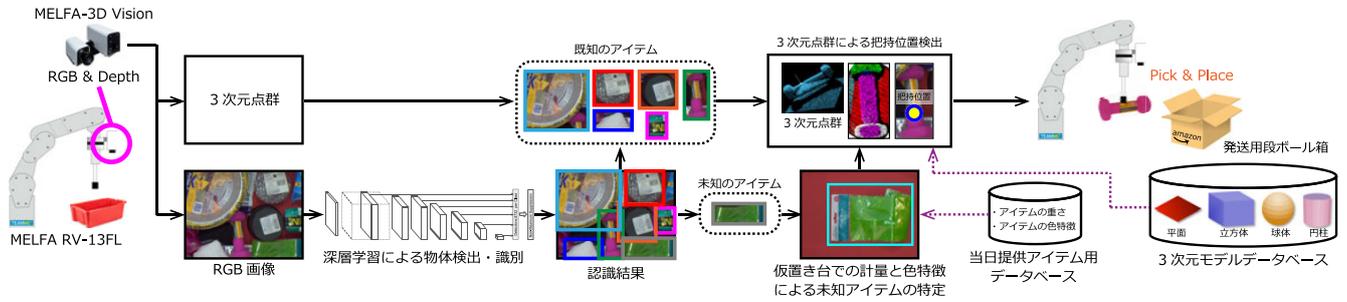


図 8 Team MC<sup>2</sup> のビジョン戦略。

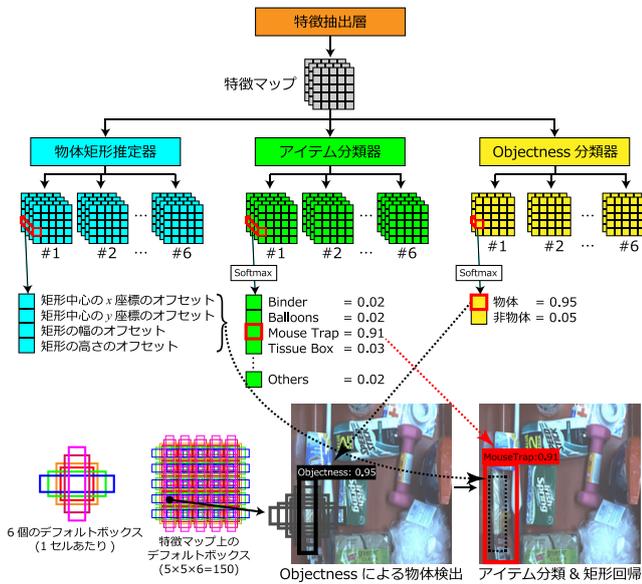


図 9 Objectness を導入した SSD による物体検出。

で物体を把持するが、競技ではメッシュカップや軽量スプーン、衣類などの吸着が困難なアイテムが存在する。吸着できないアイテムについては、Robot 2 のハンドを小型吸着ハンドからグリッパに切り替えてアイテムを把持する。

#### 4.2 ビジョン戦略

3D ビジョンセンサで取得したデータを受け取り、図 8 に示すアルゴリズムでアイテム識別を実行する。まず、storage または tote 内のばら積みアイテムを撮影し、アイテムの特定と物体矩形を出力する。この画像認識アルゴリズムは、Single Shot MultiBox Detector (SSD) [7] をベースとして開発した。SSD は、Deep Convolutional Neural Network による物体認識法であり、特徴抽出層から得られた特徴マップを入力とするアイテム分類器と物体矩形推定器から成る。しかし、このネットワーク構成の場合、大量の画像データで学習した事前提供アイテムは高精度に検出できるが、当日提供アイテムは検出することができない。そこで、従来の SSD に物体らしさを示す“objectness”分類器を導入し、未知アイテムを検出する。図 9 に objectness を導入した SSD による物体検出例を示す。Objectness 分類器は物体または非物体の 2 クラスのみを分類する畳み込み層であり、物体らしさのスコアが高い矩形を全て検出する。アイテム分類器は事前提供アイテム + それ以外 (others) の 41 クラスを分類する畳み込み層であり、検出した矩形のアイテムを特定

する。既知アイテムのクラススコアが高い場合は、その結果をそのまま採用し、others のスコアが高い場合は未知アイテムとして結果を出力する。未知アイテムとして認識した物体に対しては、仮置き台へ移動させ、アイテム重量による絞り込みを行う。さらに、未知アイテムの色特徴量を計算し、当日提供アイテム用データベースの色特徴量とマッチングさせることで未知アイテムが何のアイテムであるかを特定する。競技開始前の 30 分間に当日提供アイテムの多視点画像の撮影と計量を行うことで、競技開始までに当日提供アイテム用データベースを生成する。最後に、識別したアイテムの矩形内の 3 次元点群を使用して把持位置を検出する。把持位置の検出は、あらかじめ平面や円柱などのシンプルな 3 次元モデルをデータベースとして保持しておき、物体矩形内の 3 次元点群とデータベース内のモデルをフィッティングさせる。データベースの中で最もフィットする 3 次元モデルを算出することで、矩形内のアイテムの姿勢を近似的に求めることができる。姿勢・距離・面形状の情報に基づいて、最も把持のしやすい位置を決定する。

#### 4.3 特徴

Team MC<sup>2</sup> の特徴は、2 台のロボットアームによる効率的なピッキングである。1 台目のロボットがアイテムを把持して、仮置き台や段ボール箱に移動させている最中に同時並行で 2 台目のロボットが storage や tote にアクセスし、認識処理を実行する効率的なピック&プレースシステムを実現した。Robot 2 には、吸着型とグリッパ型のハンドを搭載しており、競技前半では吸着可能なアイテムを優先して把持し、競技後半ではハンドをグリッパ型に切り替え、吸着困難なアイテムを把持する。

また、仮置き台の設置もチームの特徴の 1 つである。認識処理において、信頼度の低い認識アイテムや未知アイテムは仮置き台へ移動させてから再度認識処理を実行する。仮置き台には重量計が搭載されており、アイテムの重さを計測することができる。さらに、仮置き台の上では基本的にアイテムが単体となるため、物体の画像特徴を抽出しやすく、アイテム同士のオクルージョンが発生しない。そのため、仮置き台を使用することで信頼度の低い認識アイテムや未知アイテムの認識を単純化し、誤認識を低減させることができる。

### 5. チーム紹介：Team K

東京大学の Team K は研究用の双腕ロボットプラットフォームを用いたシステムを構築した。以下に、Team K のシステム

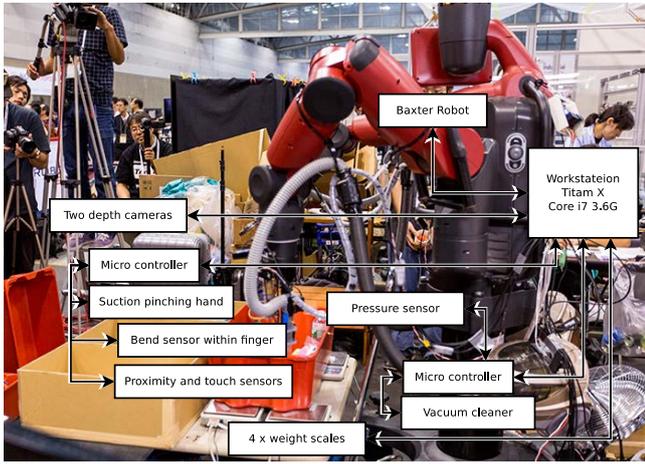


図 10 Team K のロボットシステム全体像。

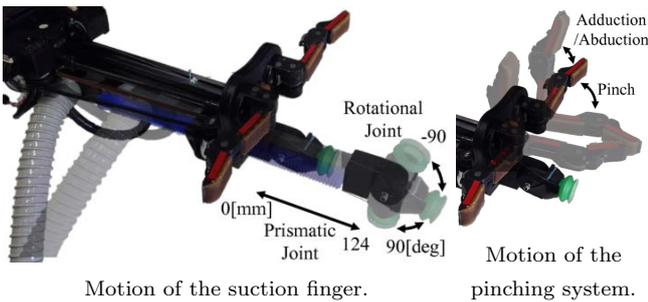


図 11 Suction Pinching Hand の構造。

構成, ビジョン戦略, 並びに特徴について述べる。

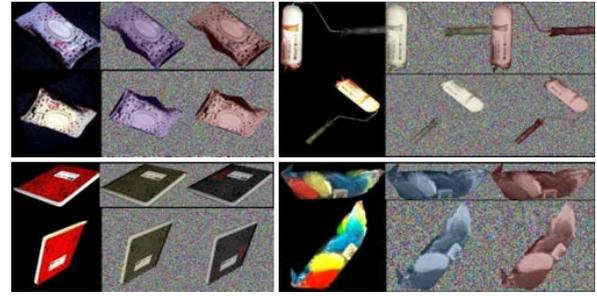
### 5.1 システム構成

Team K のロボットシステムを図 10 に示す。双腕ロボット (Rethink Robotics 社 Baxter Research Robot) に、伸縮と向きを変えられる 2 自由度の 1 本の吸引指と、2 自由度 1 駆動で 2 本の指の向きが同期して変えられる摘み指の合計 3 指をもつ Suction Pinching Hand [8] を取り付けている。図 11 はこのグリッパの構造図である。センサシステムとして、各摘み指には曲げセンサ (Spectrasymbol 社製薄膜型変位センサ FS-L-0095-103-ST) と近接触覚センサ (Robotic Materials 社製) を持ち、視覚センサとしては各腕に距離センサ (Orbbec 社製 AstraMini) をそれぞれ 2 つ設置している。さらに物品箱 (tote) の下には 4 つの電子天秤 (A&E EW-12KI) を設置した。また、吸引には掃除機を利用し、その途中に気圧センサ (Bosch 社 BME270) を設置した。

### 5.2 ビジョン戦略

ARC 2017 の特徴として直前に与えられる未知物品への対応がある。昨年度は棚に取められた物品を写した大量の画像を取得し、人によるアノテーションにより物品毎にラベルを付与したラベル画像を生成した。これをディープラーニングの学習データとすることで認識器を構成した。

一方、本年度は人によるアノテーションによるラベル画像を生成する時間が無いことから、少数の物体情報から物品ラベル画像を人工的に生成し、これを学習用データとする Few-shot Learning 手法を開発した。



Data Augmentation の結果の例

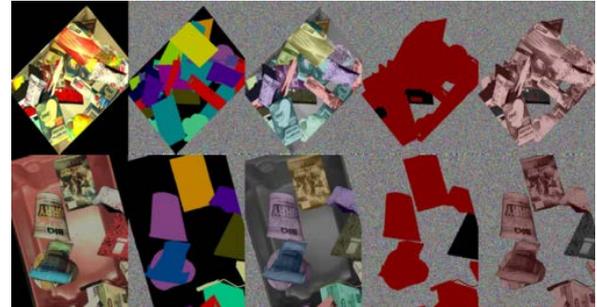


Image Stacking の結果の例

図 12 人工アノテーション画像生成。

まず、事前に配布された 40 個の既知の物品に対しては 175 枚の画像を取得し、そこから約 1500 個の物体ラベルを手動で生成している。この学習用データを用いて FCN を構成した。

一方、大会中に配られた 16 個の物品に対しては、各々 6 枚の画像が与えられている。そこで、まず、Data Augmentation として RGB 画像を HSV 画像に変換し、彩度 (Saturation) のランダムなスケール、ガウスノイズの付与、ならびに、位置、拡大縮小、上下左右反転、Shear 変換を行い、次にこれらの画像を用いて Image Stacking を行う。これは ARC で出現する背景画像 (tote とダンボール箱) に対し、生成した物品画像をランダムな位置に配置し、画像と対応する物体ラベル画像を生成する。生成した人工画像の例を図 12 に示す。

1 秒間に 2.5 回処理を行い 30 分で 4,000 枚の人工画像を生成した。背景画像としては既知物体の映った画像、及び映っていない画像の二種類を用いた各画像には 5 個から 25 個の平均 25 個の物品が映っており、合計 100,000 個の物体ラベルを自動で生成した。

大会では Titan X GPU (pascal) 1 枚を挿した Intel (R) Core (TM) i7-6850K CPU (3.6GHz) の計算機を用いて処理を行い、Data Augmentation と Image Stacking を用いた画像の生成にかかる計算時間は約 0.4 秒/枚 (2.5Hz) であり、またネットワークの再学習にかかる時間は約 0.2 秒/枚 (5Hz) であった。

人工生成したアノテーション画像と人によるアノテーション画像を用いた Semantic Segmentation 学習の比較評価を表 3 に示す。学習には最大 50,000 回の反復とし Adam (adaptive moment estimation) を用いて最適化を行った。APC 2016 では全ての物品が既知であり事前に与えられていたが、約 600 アノテーションで 40 物品の認識に約 50 IU、ターゲットとした棚の中の物品の事前知識を用いて対象物を 3-5 物品に絞れば

表 3 人工生成したアノテーション画像と人によるアノテーション画像を用いた Semantic Segmentation 学習の比較評価.

Dataset	pixel acc.	mean acc.	mean IU	fw. IU
human ( $N(m) : 40$ )	91.2	83.1	72.5	84.1
auto ( $N(m) : 40$ )	73.2	51.3	32.3	62.9
auto -aug. ( $N(m) : 40$ )	72.5	42.1	28.4	63.5
auto ( $N(m) : 16$ )	85.6	67.8	43.3	80.0
auto -aug. ( $N(m) : 16$ )	85.7	58.8	36.4	79.8

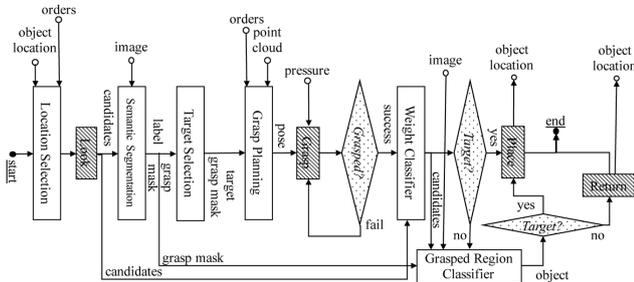


図 13 Team K の把持処理パイプライン.

70 IU となっていた. すなわち, これまでの経験からは平均 IU (Intersection Over Union) が 40 から 60 程度あればタスクの遂行には十分であることがわかってきている.

比較評価表からは, 棚を 3 分割した場合の棚 1 つあたりの平均未知物体数 5.3 (=16/3) では, 人工生成したアノテーション画像を用いた場合も新規物体に対しても 50 IU を達成しており, タスクの遂行には十分な性能を有しているが, 40 程度の未知物品の場合は必要な性能には届いていないことがわかる.

また, 結果をより詳細に調べてみると, DVD や本などの剛体に関しては, 認識が成功している場合が多い一方で, 手袋などの変形する柔軟物体に対しては失敗する場合が多いことが経験的にわかっており, 今後の課題になっている.

### 5.3 特徴

ビジョン部分以外も含めた把持動作全体のパイプラインを図 13 に示す. 画像処理でシーンの Semantic Segmentation を行った後, 物品ラベルと距離画像を用いて立方体に変換し, その重心の把持を行う. 把持動作には摘み動作と, 吸引動作の 2 つが選択可能であるが, これは予め指定された把持戦略情報を用いて選択する.

把持動作には対象へのアプローチ動作と物体の摘み動作または吸引動作の 2 つの段階がある. アプローチ動作時には曲げセンサ, 近接センサ, 気圧センサの反応があれば対象物以外の物品や床と接触したと判断することとした. また, 摘み動作または吸引動作の成否もそれぞれ近接センサ, 気圧センサ, あるいは気圧センサで判定している. さらに, 物品を取り出した後に, 物品箱の下において天秤の値を計測し, 予め作成した物品質量情報と比較することで, 指定された物品が把持されているかを検証している.

## 6. チーム紹介: Team T2

Team T2 は, 鳥取大学と東芝の合同チームである. 鳥取大学のメンバーは, メディア理解研究室の岩井・西山先生と所属

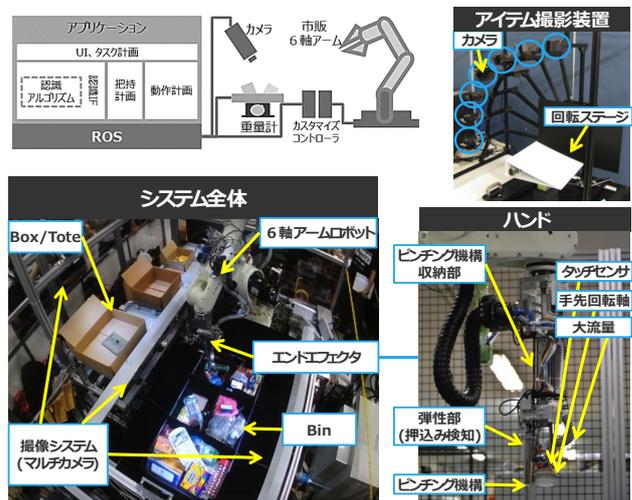


図 14 Team T2 のロボットシステム.

の学生である. 東芝 (現東芝インフラシステムズ所属を含む) のメンバーは, 様々な自動化システムの製品化を行ってきた技術者及びロボット・画像処理の研究者である. 以下に, Team T2 のシステム構成, ビジョン戦略, 並びに特徴について述べる.

### 6.1 システム構成

Team T2 のロボットシステムは図 14 のようになっている. 図中の左上はシステム構成の概略, 左下はシステム全体の俯瞰, 右下はロボットハンド, 右上は大会会場で用いたアイテム撮影装置である. システムは, 川崎重工製 6 軸アーム (RS020N), 真空ポンプ, storage を設置する台, tote および段ボール箱を設置する台から構成され, 周囲には安全柵を設置した. 台には, 各 bin・tote・段ボール箱を個別に撮影可能な RGBD カメラ (Intel®RealSense™Camera SR300) が計 8 台取り付けられている. 他のセンサ類としては, 吸着を確認するための圧力センサ, storage・tote・段ボール箱内の重さを計測する重量計がある. ハンドは, 吸着と狭持 (ピンチング) を使い分けられる. さらに, 吸着パットは 7 軸目を持ち 180 度角度を変更可能で, 狭持部は使用時以外に収納する機能を持っている. Storage は, 平置き型で 5 つの bin (大きめ 1 つ・小さめ 4 つ) で, 外周の縁を低く・透明素材を採用することでカメラの死角を減らしている. さらに, ストレージの内側の素材に IR 光を吸収するものを採用し, アイテム以外の depth ができるだけ観測されないようにしている. また, 底面に金属のメッシュを用いることで, 誤吸着が発生しにくい工夫を施した. アイテム撮影装置は, 競技時に追加される未知アイテム認識用のデータ作成に用いた. ソフトウェアは, 各コンポーネントが ROS を介して通信する構成とし, pick・stow とともに基本的な処理の流れは同じとした. 処理の流れは, bin / tote をカメラで撮影し, アイテムを識別・検出する. 同時に, 段ボール / bin を撮影し, アイテムを置く場所を調査する. 次に, アイテムへの軌道を算出し, 把持する. 同時に, アイテムを置くための位置を求めておく. 最後に, アイテムを把持した位置から置く場所までの軌道を算出し, アイテムを移動させる. 以上の処理をオーダーが終わるまで繰り返す. より詳細なシステム構成に関して

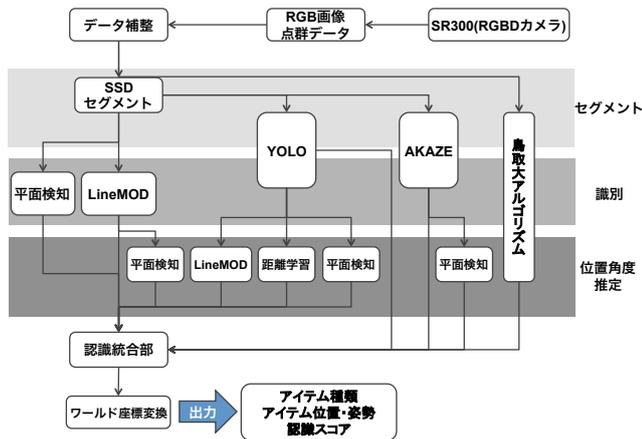


図 15 Team T2 のビジョン戦略.

は文献 [9] を参照していただきたい.

## 6.2 ビジョン戦略

Team T2 のビジョン処理の流れを図 15 に示す. RGBD を入力として, アイテム領域の検出 (セグメント), セグメント領域からのアイテム種類の識別, アイテム位置・姿勢の推定で行う. Team T2 では複数の手法を組み合わせた複数のパスで処理をしているため, 最後に複数パスからのそれぞれの結果を統合し, ビジョンの出力としている. 採用した手法は, SSD を利用したセグメント, LineMOD を用いた識別・姿勢推定, YOLO v2 を用いた検出・識別, AKAZE を用いた検出・識別, 距離学習手法による姿勢推定, 鳥取大学が開発した検出・識別手法などがある [7], [10]~[14]. これら以外にも, ロボットが把持できそうな平面だけを検出する方法も採用した. これらの組み合わせ方法は図 15 に示す通りで合計 10 種類のパスがある. このように複数のパスを利用することで, 様々な性質をもつアイテムに広く対応することが可能となった. 一方で, 最大 10 種の結果を統合する作業が必要となってしまう. 今回 Team T2 では, 事前に各アイテム・アイテムの性質 (剛体 / 変形物体・透明 / 不透明) ごとに, 手法の優先度を用意しておく方法を採用した. この方法である程度の調整は可能であるが, そのバランスをとるのは非常に難しい問題であることを感じた. 学習用データは, SR300 で各アイテムを全周から約 700 枚撮影したものをを用いた. また, ロボットシステムで実際に撮影した画像も人手によるアノテーションを行い利用した. 実機での撮影画像は 1000 枚程度であった. 一方, 鳥取大学の手法は, 訓練サンプル収集時において, 少ない撮影枚数で既存手法と同等の精度が実現できる物体検出方法 [14] である. この手法では, 検出に有効な物体側面のみを訓練サンプルとして識別器を学習させる. 撮影枚数を 1/175 回に減らしても同等の精度が実現できることを実験で確認している. 未知アイテムについては, 図 14 の右上にある撮影装置を用いて, 30 分のうちに撮影・学習を行った. 撮影装置は 1 回の操作でアイテムの片面 (表裏いずれか) の全周 168 視点からの撮影を約 45 秒で行うことができる. また, 学習時間も非常に限られたものであることから, 未知アイテムに対しては, LineMOD, AKAZE のみ学習を行い, それらを用いるパスのみで認識している.

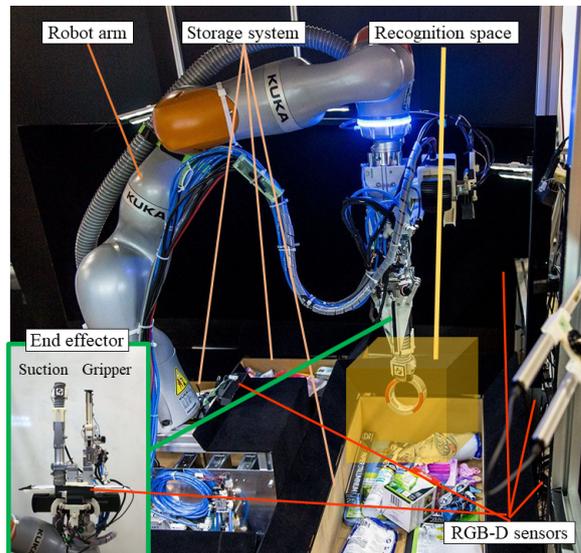


Fig. 16 Setup of team NAIST-Panasonic.

## 6.3 特徴

Team T2 の特徴として, 図 14 に示したような複数の固定カメラによる並列処理がある. それぞれの bin・tote・段ボール箱に専用のカメラを用いることで, アイテム認識やアイテムを置く場所の情報を多く取得することが可能となる. これは, アイテムの移動プランニングを立てる際に, 多くの候補から選択できるようになるため, 効率のよいプランニングが選ばれやすくなるといった効果がある. さらに, ピック対象アイテムが認識できない場合でも, 優先度の高い bin (ピック対象アイテムが一番多く隠れている bin などのルールを設定) を選択して, 邪魔なアイテムを退かす軌道計算をするという効果もある. また, 把持したアイテムを bin / 段ボール箱に効率的に置く “箱詰め計画” を計算することも特徴である. 箱詰め計画は, カメラから空きスペースの情報を取得し, アイテムの置く位置・向きなど決めている. Team T2 のロボットハンドは図 14 に示すように, 吸着と狭持をすることができる. 吸着と狭持の切り替えは, アイテム種類だけでなく, アイテムの状態でも行う. 例えば, メッシュカップの底面が検出された場合は吸着, 緑が検出された場合は狭持を選択する. さらに, 吸着はアイテムによって吸引力を変化させることができるのも特徴の一つである. これは, アイテム表面が脆い場合に, 破損してしまうのを防ぐ効果がある. また, 弱い吸着でアイテムを運ぶときには, アーム速度を低下させ, 落下を防ぐような処置もとっている. 最後に, Team T2 は競技スペースの周囲に安全柵を設置した. われわれのロボットは, 本大会で最大のサイズのものであったこともあるが, 非常に安全に注力していることにも注目してほしい.

## 7. Team Introduction : NAIST-Panasonic

The Nara Institute of Science and Technology (NAIST) and Panasonic Corporation teamed up to participate in the ARC 2017 as team NAIST-Panasonic and obtained the 6th place. With a total of 20 members, the team includes a postdoctoral researcher, Ph.D. candidates, Masters students

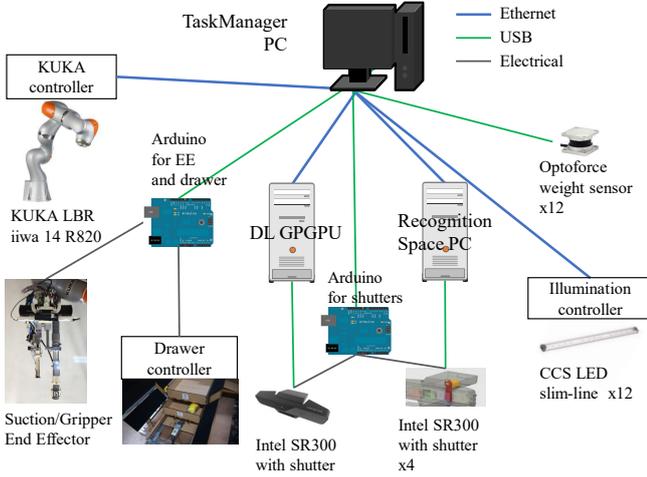


Fig. 17 System overview of team NAIST-Panasonic.

and electromechanical engineers. Previously, team NAIST had taken the 1st place at the Airbus Shopfloor Challenge 2016, an international robotics competition at ICRA 2016 [15]. The following sections summarize the developed system, vision strategy and main features of the team NAIST-Panasonic.

### 7.1 System Overview

The robotic system developed by the team NAIST-Panasonic consists of a custom-made end effector mounted on a 7-DOF serial robotic manipulator (KUKA LBR iiwa 14 R820), a controlled recognition space formed by an array of RGB-D cameras, and a shelf or *storage system* with weight sensors [16]. The setup of the developed solution is shown in Fig. 16 and an overview of the components of the system is shown in Fig. 17.

The end effector features suction and gripper tools mounted on separate linear actuators, as well as an RGB-D sensor for object recognition and grasping point estimation. The suction tool consists of a compliant, partially constrained vacuum cleaner hose, whose tip has a force-sensitive resistor (FSR) to detect contact with the items. We power the suction tool with an industrial-grade blower with a maximum vacuum of  $-40$  kPa [17]. A pressure sensor detects suction and dropped items, while a waste gate regulates the static pressure at the suction cup to avoid damaging delicate packaging. The gripper tool consists of a parallel gripper with fingers covered by a high-friction rubber. FSRs are installed between the rubber and the fingers to detect if an item has been grasped, as well as in protruding tips to detect contact with items. We use the suction tool for about 80% of the items and switch to the gripper after suction fails.

The storage system consists of three bins, one of which is a drawer whose purpose is to increase the available surface to place items. We installed two sets of four 3D force sensors (Optoforce OMD-20-FG-100N) attached to a rect-

angular base made of aluminium frames under the storage system to measure the weight of the items inside.

The recognition space consists of an array of four RGB-D cameras (Intel Realsense SR300) that observes a space above the storage system (without physical structure), where illumination is controlled with eight LEDs and non-reflective black plates control the background of the cameras' images. The purpose of this recognition space is to move the grasped items inside a partially structured environment where higher successful recognition rates can be expected.

The system is implemented in ROS Kinetic and the source code is public in the Warehouse Picking Automation Challenges repository at GitHub.

### 7.2 Vision Strategy

The strategy of team NAIST-Panasonic to recognize items consists of two steps:

- i) Look into a container with the end effector's camera, find the grasping points and items' classes from RGB images, and pick an item.
- ii) Bring the grasped item into the recognition space where multiple views of the item are used for confirmation or reclassification.

In step (i), we use a retrained version of the real-time object detection system YOLO v2 [11]. Since it is necessary to recognize items in realistic conditions, we trained YOLO using 8,000 images taken in overlapped and crowded conditions and manually annotated (class, bounding box and grasping points). Additionally, we trained it with 50,000 images generated by randomly inserting pictures of items without background into pictures of empty containers and automatically annotated. This deep learning approach was used for the *known* items (*i.e.*, items available months before the challenge and which would become half of the items at the live competition).

In step (ii), we use features such as color histogram and bounding box volume calculated from four RGB-D cameras inside an environment where controlled illumination and background facilitate the object recognition. Robust recognition can be achieved using SVMs for single or combined features (*e.g.*, volume and weight), and even Euclidian distance yields reasonable results. Since feature values can be acquired easily, this approach is applicable to the recognition of *unknown* items, which are released for 30 minutes shortly before the round. In fact, data can be collected in about 90 seconds for each unknown item.

Finally, the results of each recognition method are summarized in a weighted sum of confidences, where the highest overall confidence determines the item class.

### 7.3 Main Features

By combining learned and manual features into a 2-step

recognition process, we take advantage of both approaches and achieve a robust recognition for known and unknown items. In particular, coupling the bounding box volume and weight yielded effective recognition rates even for deformable objects and clamshell-type items. Moreover, we decreased the risk of losing items and increased the confidence of the recognition results by using a controlled recognition space where no re-grasping was necessary.

The smart design of the end effector increased the reliability and consistency of the system. Especially, the difficulty of system requirements such as path planning was eased due to the high flow rate and the compliance of the suction tool. Furthermore, the robot was able to safely transport the items thanks to the high suction power.

The system can continue the tasks even when something fails, as failures are considered part of the system operation and alternative behaviors are prepared in advance to overcome the errors. Our design philosophy was to keep the system running and scoring points: from modern development tools and sanity checks to drawer shaking and desperate modes to pick items, our system was prepared.

## 8. ARC の課題と今後

Amazon Robotics Challenge は、2017 年をもって競技会が今後開催されないことが発表された。3 年間に亘る大会を通じて、深層学習の進展とともにピッキング能力が大幅に向上したが、あくまでも大会という枠の中での進展である。実問題を解くには、まだマニピュレーションの基礎的な問題を解く必要があり、今後は基礎研究に力を入れて行くという方針である。

ARC の参加を通じて、実際にロボットシステムを構築し、その性能を評価し、競技会を通じて技術的な課題を分析し議論することで、新たな問題を知ることができたことは大きな収穫であったと思う。最後に主催者である Amazon Robotics の Joey Durham 氏による ARC における Research questions をここにシェアしておく。

- (1) How do you recognize an object after seeing it only once?
- (2) How can you learn to pick up a huge variety of items?
- (3) How do you create and then pick from densely packed items?
- (4) How do you adapt when things don't go to plan?

### 文 献

- [1] R. D'Andrea, "Guest editorial: A revolution in the warehouse: A retrospective on kiva systems and the grand challenges ahead," *Automation Science and Engineering*, vol.9, no.4, pp.638–639, 2012.
- [2] N. Correll, K.E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J.M. Romano, and P.R. Wurman, "Analysis and Observations From the First Amazon Picking Challenge," *Automation Science and Engineering*, vol.15, no.1, pp.172–188, 2018.
- [3] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F.R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, et al., "Robotic Pick-and-Place of Novel Objects in Clutter with Multi-Affordance Grasping and Cross-Domain Image Matching," *arXiv preprint arXiv:1710.01330*, 2017.
- [4] H. Fujiyoshi, T. Yamashita, Y. Yamauchi, R. Murata, T. Hasegawa, M. Kaneko, Y. Murai, M. Hashimoto, S. Akizuki, M. Nagase, Y. Sakuramoto, S. Takei, S. Itoh, Y. Domae, R. Kawanishi, K. Shiratsuchi, R. Haraguchi, and M. Fujita, "Combined Point Cloud and Appearance-Based Object Detection for Grasping Rigid and Non-Rigid Objects," *International Workshop on Recovering 6D Object Pose at ICCV*, 2015.
- [5] H. Fujiyoshi, T. Yamashita, Y. Yamauchi, T. Hasegawa, M. Hashimoto, S. Akizuki, Y. Domae, and R. Kawanishi, "Team C<sup>2</sup>M: Two Cooperative Robots for Picking and Stowing in Amazon Picking Challenge 2016," *Warehouse Picking Automation Workshop at ICRA*, 2017.
- [6] 藤吉弘亘, 松元叡一, 岡田慧, "[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*, pp.123–129, 2017.
- [7] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision*, pp.21–37, 2016.
- [8] S. Hasegawa, K. Wada, Y. Niitani, K. Okada, and M. Inaba, "A Three-Fingered Hand with a Suction Gripping System for Picking Various Objects in Cluttered Narrow Space," *International Conference on Robotics and Systems*, pp.1164–1171, 2017.
- [9] 江原浩二, 野口弘貴, 小川昭人, "アマゾンロボティクスチャレンジを通じたロボットプラットフォームの開発," *東芝レビュー*, 2018.
- [10] S. Hinterstoisser, S. Holzer, C. Cagniard, S. Ilic, K. Konolige, N. Navab, and V. Lepetit, "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes," *International Conference on Computer Vision*, pp.858–865, 2011.
- [11] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Conference on Computer Vision and Pattern Recognition*, pp.7263–7271, 2017.
- [12] F.A. Pablo, B. Adrien, and J.D. Andrew, "KAZE Features," *European Conference on Computer Vision*, pp.214–227, 2012.
- [13] P. Wohlhart and V. Lepetit, "Learning descriptors for object recognition and 3D pose estimation," *Conference on Computer Vision and Pattern Recognition*, pp.3109–3118, 2015.
- [14] 上野高貴, 西山正志, 岩井儀雄, "物体側面を用いた訓練サンプルによる検出手法の検討," *ビジョン技術の実利用ワークショップ, OS3-H4*, pp.292–299, 2017.
- [15] G.A. Garcia Ricardez, F. von Drigalski, L. El Hafi, M. Ding, J. Takamatsu, and T. Ogasawara, "Lessons from the Airbus Shopfloor Challenge 2016 and the Amazon Robotics Challenge 2017," *System Integration Division Annual Conference*, pp.572–575, 2017.
- [16] G.A. Garcia Ricardez, F. von Drigalski, L. El Hafi, S. Okada, P. Yang, W. Yamazaki, V. Hoerig, A. Delmotte, A. Yuguchi, M. Gall, C. Shioyama, K. Toyoshima, P.M. Uriguen Eljuri, R. Elizalde Zapata, M. Ding, J. Takamatsu, and T. Ogasawara, "Warehouse Picking Automation System with Learning- and Feature-based Object Recognition and Grasping Point Estimation," *System Integration Division Annual Conference*, pp.2249–2253, 2017.
- [17] G.A. Garcia Ricardez, L. El Hafi, F. von Drigalski, R. Elizalde Zapata, C. Shioyama, K. Toyoshima, P.M. Uriguen Eljuri, M. Gall, A. Yuguchi, A. Delmotte, V. Hoerig, W. Yamazaki, S. Okada, Y. Kato, R. Futakuchi, K. Inoue, K. Asai, Y. Okazaki, M. Yamamoto, M. Ding, J. Takamatsu, and T. Ogasawara, "Climbing on Giants Shoulders: Newcomers Road into the Amazon Robotics Challenge 2017," *Warehouse Picking Automation Workshop at ICRA*, 2017.