社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[特別講演] Amazon Picking Challenge 2016 の参加レポート

藤吉　弘亘†　　松元　叡一††　　岡田　　慧†††

† 中部大学　〒 487–8501 愛知県春日井市松本町 1200
†† 株式会社 Preferred Networks　〒 100–0004 東京都千代田区 大手町 1 丁目 6–1 大手町ビル 2F
††† 東京大学　〒 113–8656 東京都文京区本郷 7–3–1
E-mail: †hf@cs.chubu.ac.jp, ††matsumoto@preferred.jp, †††k-okada@jsk.t.u-tokyo.ac.jp

あらまし　e コマースにおける倉庫では，棚内の多品種商品を識別し，把持計画により安定したピックアンドプレースを実現することが自動化に向けた重要課題である．Amazon.com, Inc. が 2015 年に開催したロボット国際競技大会である Amazon Picking Challenge では，まさにこの多品種のピッキング問題が取り上げられた．本講演では，Amazon Picking Challenge 2016 に参加した各チームの取り組みを紹介し，そこから分析した今後の技術課題についても議論する．

キーワード　Amazon Picking Challenge，ピッキング，ロボット競技，深層学習

1. はじめに

米国 Amazon.com の物流倉庫では，kiva systems (2016 年現在 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ピッキング担当者まで自動搬送する．現在，棚からのピッキングにおいては人手によって行われている状況であり，ピッキングロボットによる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内に無作為に収められているため，棚内の多品種商品を識別し，把持計画により安定したピックアンドプレースを実現することが，自動化に向けた最重要課題である．このような背景の下，米 Amazon.com は，2015 年 5 月 26~28 日に米国・シアトルにて物流の自動化技術を競う初のロボット大会「Amazon Picking Challenge」を開催した．第 1 回となる Amazon Picking Challenge (APC) 2015 では，多品種のピッキング問題が取り上げられ，棚の Bin と呼ばれる 12 個の枠の中から指定された 25 種類の商品 (アイテム) を取り出すピッキングタスクが競技対象であった [1]．ピッキング対象のアイテムの種類には，剛体，非剛体，半透明物体で構成されており，単一の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．2016 年 7 月 1~2 日にドイツ・ライプツィヒにて開催された第 2 回目となる APC 2016 では，競技内容をより実問題に近づけるため，Pick task と Stow task の二つの競技タスクが設定された．また，対象となるアイテムも 25 種類から 39 種類へと増加した．本稿では，Amazon Picking Challenge 2016 に参加した各チームの取り組みを紹介し，そこから分析した今後の技術課題についても議論する．

2. APC 2016 の競技タスク

APC2016 の Pick task と Stow task において，対象となるアイテムを図 1 に示す．以下では，各タスクの概要と評価方法について述べる．

図 1　APC 2016 での対象アイテム．

2.1 Pick task

Pick task は，12 個の Bin から指定されたアイテムを一つずつ，計 12 個をピッキングするタスクである．APC2015 の競技内容と同様であるが，アイテム数の増加や重なりの発生など，実利用に向けてより難易度が高くなった．39 種類 46 個のアイテムが各 Bin に最大 10 個置かれ，Bin 内ではアイテム同士の

図 2　Pick task のアイテム配置例.

図 3　Stow task のアイテム配置例.

接触や重なりがあるため，アイテムの一部に隠れが発生している状態である．また，同一アイテムが複数存在する場合もある．図 2 に Pick task のアイテム配置例を示す．このようにアイテムが配置された Bin の中から指定されたアイテムをピッキングするには，物体認識，三次元処理，物体把持計画等を統合するロボット技術が要求される．

2.2　Stow task

Stow task は，APC 2016 から始まった新しいタスクである．Tote と呼ばれる箱に無造作に置かれた 12 個のアイテムをピッキングし，棚の Bin に収納するタスクである．各 Bin には 1~10 個のアイテムが既に置かれているため，物体認識，三次元処理，物体把持計画に加え，それらとの衝突判定を考慮しながら収納する技術が必要である．図 3 に Stow task のアイテム配置例を示す．

両タスクの競技前には，主催者が ±3cm の範囲内で棚の位置を故意にずらす．このため，競技開始後にロボットは棚との位置関係を自動でキャリブレーションする必要がある．これは実際の Amazon の倉庫を想定し，棚を運ぶ自律ロボット Kiva Pod の自己位置の精度に合わせたものと想定される．

2.3　評 価 方 法

競技時間は Pick task と Stow task ともに 15 分であり，1 アイテムの Pick もしくは Stow が成功すると，表 1 のように Bin の中のアイテムの個数 (難易度) に応じて得点が変化し，アイテム数が多いほど高得点となる．なお，アイテム毎には，認識や把持の難度に応じた 1~3 点のボーナスポイントが設定されている．一方，アイテムを 30cm 以上の高さから落下させたり，破損させた場合には減点となる．いずれのタスクにおいても，計 12 個の指定されたアイテムを対象とし，15 分間の合計ポイント数を競う．また，競技途中でリセットする場合には，残り時間から 1 分の競技時間削減のペナルティが課せられ，競技を始めから開始する．

表 1　Bin の中のアイテム数と得点

Bin の中のアイテム数	得点
1~2 アイテム	10 ポイント
3~4 アイテム	15 ポイント
5 アイテム以上	20 ポイント

3.　APC 2016 の競技結果

APC 2016 では事前審査を通過した 16 チームが競技会に出場した．事前審査では，各チームの技術力を示すロボットのデモンストレーション動画とインタビューが行われ，16 チームがファイナリストとして選ばれた．米国からは 5 チーム，日本からは 4 チームの参加であった．APC2015 では日本からの参加は 2 チームのみであったが，APC 2016 では 4 チームに増え，日本における APC の注目度の高さを表している．他の参加国は，ドイツ，オランダ，オーストリア，スウェーデン，インドであった．表 2 に，全参加チームの所属と使用しているロボットを示す．双腕ロボットの Baxter が 6 台と多く，後述する高順位のチームは産業用ロボットを使用する傾向があった．エンドエフェクタには，吸着タイプと挟持タイプの両者を装備し，アイテムに合わせて使い分けるロボットが多く見られた．

3.1　Stow task の競技結果

大会初日 (7 月 1 日) は，Stow task の競技が開催され，1 位はドイツの Delft (214 ポイント)，2 位はドイツの NimbRo Picking (186 点)，3 位は米国の MIT (164 点) であった．1 位の Delft は，図 4 に示すように Tote 専用の上部に固定したカメラから Tote 内のアイテムを認識し，単腕ロボットの細長い吸着ハンドでアイテムを把持する．棚に移動して収納する際には，収納する方向に対して吸着ハンドの向きを下向きに可動し，棚にアイテムを収納する．Tote 内の物体認識は，ロボットがストーイング動作の間に固定カメラで行うため，全体の時間を短縮している．競技では，11 個全てのアイテムのストーイングを約 7 分 30 秒で完了した．なお，Tote 内の物体認識には Deep Learning が使用された．

3.2　Pick task の競技結果

翌日 (7 月 2 日) の Pick task の競技では，1 位は Stow task でも 1 位となった Delft (105 点)，2 位は PFN (105 点)，3 位は NimbRo Picking (97 点) であった．1 位の Delft と 2 位の PFN は同得点であったため，ルールに従い，一つ目のアイテムのピッキングに成功したまでの時間が短い Delft が 1 位となった．Delft のエンドエフェクタは，図 5 に示すように可動型吸着と挟持ハンドが合体した作りとなっており，対象アイテム毎に把持方法を切り替えていた．一つのエンドエフェクタに吸着と挟持の両者の機能を持ち，かつできるだけ細くした機構は，多種多様なアイテムを衝突せずにピッキングするのに適しており，素晴らしいアイディアである．他のチームが苦労していた吸着では把持が難しい網状のペン立てを，ペン立ての縁をつま

表 2　APC 2016 参加チーム一覧 (アルファベット順)

チーム名	所属	国	ロボット	ハンドの種類
AA-team	The University of Tokyo	日本	Seed solutions (双腕)	挟持型
ACRV	Queensland University of Technology & University of Adelaide	オーストラリア	Baxter (双腕)	吸着型
Applied Robotics	Smart Robotics bv, KU Leuven, and Alten Technology	ベルギー	smart robotics (単腕)	吸着型
C^2M	Chubu University, Chukyo University, and Mitsubishi Electric	日本	MELFA (単腕 ×2 台)	挟持型，吸着型
Dataspeed-Grizzly	Dataspeed Inc & Oakland University	アメリカ	Baxter (双腕)	挟持型，吸着型
Delft	TU Delft & Delft Robotics	オランダ	Yaskawa (単腕)	吸着型
Duke	Duke University	アメリカ	Baxter (双腕)	吸着型
HARP	Carnegie Mellon University	アメリカ	Barrett Technology	吸着型
IITK-TCS	Indian Institute of Technology Kanpur & Tata Consultancy Services	インド	Robotnik(単腕)	吸着型
KTH	Kungliga Tekniska Högskolan	スウェーデン	Baxter (双腕)	挟持型，吸着型
MIT	Massachusetts Institute of Technology	アメリカ	ABB (単腕)	挟持型，吸着型
NimbRo Picking	University of Bonn	ドイツ	Universal Robot (単腕)	吸着型
PFN	Preferred Networks, Inc.	日本	FUNUC (単腕 ×2 台)	挟持型，吸着型
Robological + UC_SMaRTi	Robological PTY, University of New South Wales, and University of Canberra	オーストラリア	Baxter (双腕)	吸着型
Rutgers ARM	Rutgers University & UniGripper	アメリカ	Yaskawa (単腕)	挟持型，吸着型
Team K	The University of Tokyo	日本	Baxter (双腕)	吸着型

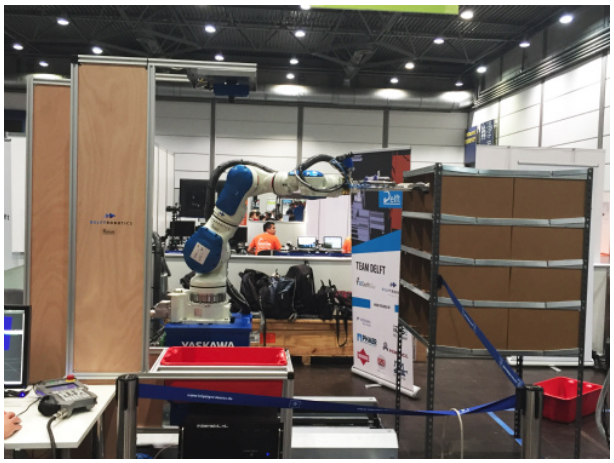

図 4　Delft (オランダ) のロボット．アーム部分に固定しているカメラから Tote 内のアイテムを認識して，細長い吸着ハンドでアイテムを把持する．

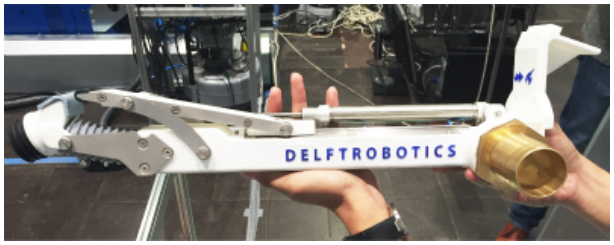

図 5　Delft のロボットのエンドエフェクタ．

むようにして把持していた．全チームの傾向としては，Stow task と比べると全体的に Pick task の得点が低い．縦 19cm×横 25cm と狭い枠の中に複数のアイテムが収納され，かつ指定アイテムが奥に配置されている場合に，ピッキングが非常に難しくなるためである．全 12 アイテムのピッキングに成功したチームは無かった．APC2015 の優勝チームは，11 アイテムのピッキングに成功したが，実利用の想定に沿ったルールの改変により，Pick task の難しさが今回浮き彫りとなったものと言える．また，Stow task と Pick task の上位はほぼ同じチームが占めており，Pick task の方が難しいタスクではあるが，これらのタスクが共通の技術で解決できることを示唆している．

3.3　APC 2016 参加ロボットの傾向

競技大会の 3 日目には，チーム毎に 15 分のデモストレーションがあり，各ロボットの特長をプレゼンテーションする機会があった．APC 2016 の上位チームの傾向として，単腕の産業ロボットと吸着ハンドによるピッキングと，Deep Learning を用いた物体認識の二つが挙げられる．エンドエフェクタとして細長い軸の先に装着された可動型の吸着ハンドを用いることで，奥に収納されたアイテムのピッキングが可能である．また，アイテムの認識 (セグメンテーションを含む) や最適な把持位置検出に Deep Learning を用いることで高精度なロボットビジョンを実現している．

4.　チーム紹介：Team C^2M

中部大学・中京大学・三菱電機の合同チームである Team C^2M は，2 台の独立した産業用ロボットアームを活用したロボットシステムを構築した．以下に，Team C^2M 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．

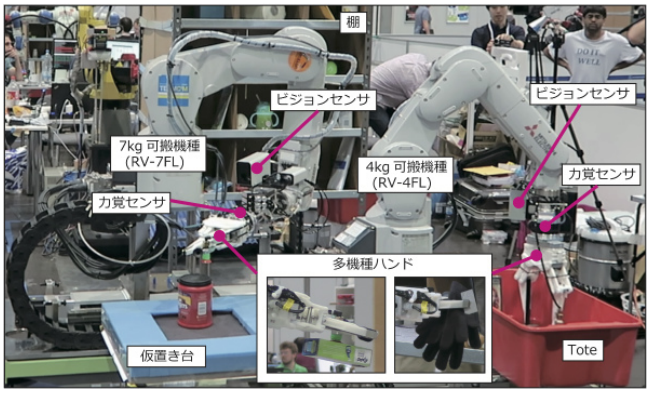

図 6　Team C^{2}M のロボットシステム全体像.

4.1　システム構成

Team C^{2}M のロボットシステムは，図 6 に示すように垂直多関節型の産業用ロボット 2 台 (7kg 可搬 MELFA RV-7FL と 4kg 可搬 MELFA RV-4FL) にそれぞれ 3 次元ビジョンセンサ (MELFA-3D Vision)，力覚センサ (4F-FS001)，多機能ハンドが搭載され，7kg 可搬のロボットが 1 軸の走行台上に設置されている．力覚センサは，棚やアイテムとの衝突判定や把持したアイテムの計量に利用する．また，アイテムを仮置きするための台が設置されている．

棚に直接アクセスするロボットには，様々な姿勢のアイテムを取るために，リーチが長く可搬重量の大きい 7kg 可搬のロボットを使用した．Tote からアイテムを取り出すロボットは必要十分なリーチと可搬重量である 4kg 可搬のロボットを用いた．3 次元ビジョンセンサは，カメラとプロジェクタからなるアクティブステレオ法で RGB 画像と距離画像を出力する．このセンサをアイテムの 3 次元計測と認識に応用する．

4.2　ビジョン戦略

3 次元ビジョンセンサで取得した計測データを受け取り，アイテムの識別を実行する．Convolutional Neural Network (CNN) ベースの認識手法により，入力画像中に存在する物体がどのアイテムであるかを同定するとともに，あらかじめ用意しておいたアイテムのデータベース画像とキーポイントマッチングする．キーポイントマッチングにおけるキーポイント検出器は Cascaded FAST detector [2]，局所特徴量記述子は ORB descriptor [3] を用いる．キーポイントマッチングでは特定のアイテムを高速に認識することが可能であるが，テクスチャの少ないアイテムに対してはマッチングが困難である．そこで，CNN ベースのアイテム認識を導入している．CNN はあらゆる視点から撮影された大量のアイテム画像を学習させる．CNN は認識精度が高い一方でフィルタの畳み込み演算が多いため，計算コストが高くなる問題がある．そこで，画像から把持可能な位置を全て検出し，検出された把持位置の周辺画像を用いて CNN でアイテムを識別する．このように，最初に把持位置を検出することで，検出された把持位置のみに認識処理を実行すれば良いため効率的なアイテム識別を実現できる．把持位置は，Fast Graspability Evaluation [4] を用いて高速に検出することが可能である．最後に仮説検証 [5] により入力シーンを解釈することで現実的に起こり得ないラベル配置が棄却され，正確なアイテムの同定が可能となる．図 7 に Team C^{2}M のビジョン戦略の流れ図を示す．

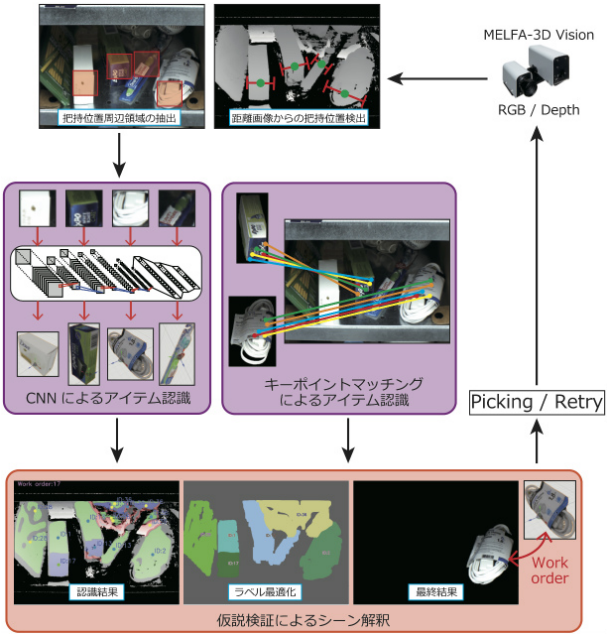

図 7　Team C^{2}M のビジョン戦略.

4.3　特　　徴

Team C^{2}M の特徴は，仮置き台を挟んだ 2 台のロボットアームの協調した作業と力覚センサの利用である．Stow task では Tote の近くに設置されたロボットアーム (RV-4FL) が，Tote と仮置き台の間で作業し，Tote から取り出したアイテムを仮置き台に移動させる．その後，棚の近くに設置されたロボットアーム (RV-7FL) が仮置き台に置かれたアイテムを認識して棚に収納する．2 台のロボットが協調することで，Stow task の問題を簡単化し，効率の良いアイテム収納を実現した．

ロボットアームの先端に取り付けられた力覚センサは，商品を把持する際にロボットや棚，アイテムに不必要な力が加わっていないかを判定することで，衝突や破損を防止することができる．また，把持したアイテムの計量により，識別結果と把持したアイテムが一致しているかを確認し，誤認識を減らすことができる．さらに，力覚センサを搭載したハンドで棚に触れることで，基準位置からの微細な位置ずれを判定し，棚とロボット間の自動キャリブレーションを実現した．

以上が Team C^{2}M のロボットの特徴であり，製品としての高い信頼性の産業用ロボット技術を中心に応用したロボットシステムである．ファクトリオートメーション (FA) の分野で培われた精度・速度を重視したシステムに，新たに多品種対応向けの技術を加えた本システムである．

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 のシステム構成，ビジョン戦略，並びに特徴について述べる．

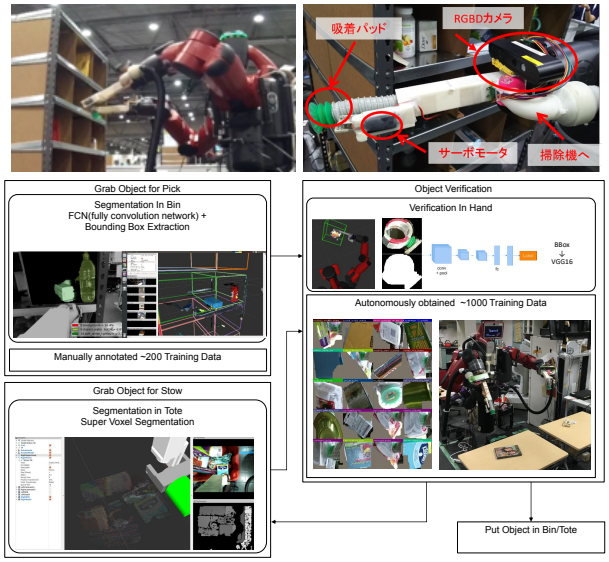

図 8　Team K のロボットシステム全体像.

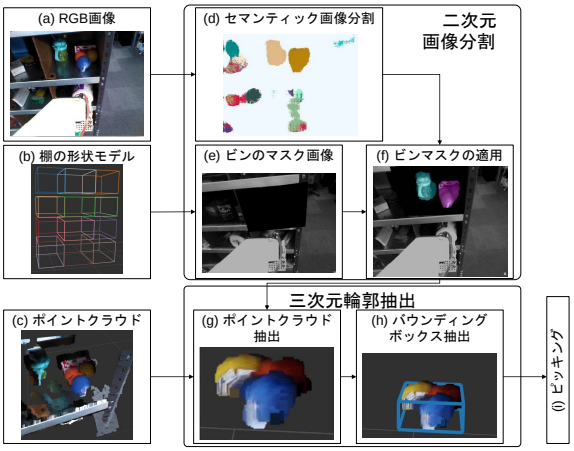

図 9　Team K のビジョン戦略.

5.1　システム構成

Team K のロボットシステムは図 8 の上図に示すように双腕ロボット (Rethink Robotics 社 Baxter) に，1 自由度のアクチュエータを有しパッドの向きを変えられる自作の吸引グリッパ，RGB-D センサを胸に 1 台 (Microsoft 社 Kinect2), 各アームに 1 台づつ (Orbbec 社 Astra S) を取り付けた構成になっている.

全体のシステム構成は図 8 の下図に示すように Pick task 用の認識行動部，Stow task 用の認識行動部に加えて，両タスクで共通に利用している認識と行動の行為検証部の 3 つのサブモジュールから構成される.

Pick task 用の認識行動部では棚の中の物品を後述のセマンティック画像分割と三次元物品輪郭抽出によりバウンディングボックスとして表現し，これをヒューリスティックな Pick 戦略で吸引把持する．一方，Stow task 用の認識行動部では入力点群を Supervoxel 法で領域分割し，その大きい領域から吸引把持していく.

行為検証部では吸引把持した物品の認識および吸引把持動作が成功したかを検証するために，吸引把持している物品をアームを移動させることで胸部に取り付けた RGB-D センサの前に移動させ，Deep Learning 画像認識により物品の識別を行い，前段の認識と行動が正しかったか，すなわち正しい部品が手先に存在するかを検証し，正しくない場合には，物品を元に戻し再度動作を実行する.

5.2　ビジョン戦略

図 9 に Pick task 用のアイテム識別と位置同定法の流れ図を示す [6]．まず，オフライン処理として物体認識のための学習データは後述の方法で予め収集しておく．オンライン処理としてはアームに取りつけた RGB-D センサ（観測レンジ 0.3-5.8[m], Structured Light IR Projection 方式) を用いて RGB 画像を取得し (a)，セマンティック画像分割によって画像内の各ピクセルがどの物品にラベルづけされるのかを決定する (d)．一方で事前に与えられた棚の形状モデル (b) を用いて，現在の RGB-D センサの位置姿勢から見た各ビンに対するマスク画像を生成し (e)，これを用いて目的のビンの内部領域だけに対応した物品ラベル情報を取得する (f)．また，RGB-D センサで得られるポイントクラウド (3 次元点群情報)(c) のうち，目的の物体ラベルを持つピクセルに対応する点群を抽出し (g), これのバウンディングボックスを用いて物品の重心と概形状を取得し (h)，Pick 戦略実行モジュールを駆動する.

セマンティック画像分割の Convolutional Neural Network (CNN) は 16 層の畳み込み層からなり，最終層として転置畳み込み層を用いた FCN-32s [7] を元に，過去のパラメータ更新の履歴を活用しながら適応的に学習率を調整していく ADAM [8] を最適化手法として構成した．また，学習済みの VGG16 net [9] モデルを用いて 16 層のうち前段の 13 層の畳み込み層の重みを初期化した．深層学習のフレームワークは Chainer [10] を用い，GPU は NVIDIA TitanX を利用した．認識検証部では VGG16 net を 40 クラスの物体識別器として用いており，学習の際には学習済みモデル [9] を元に重みを初期化し，輝度，Crop，Translation，Rotation 等のデータ増強を行なったデータセットで学習を行なった．学習のためのデータセットは図 8 の下図に示すようなデータセット収集システムにより作成した.

FCN の入力は RGB 画像で，出力は各ピクセルについて，クラスの候補それぞれが割り当てられる確からしさを求めた 3 次元配列である．クラスの候補としては物品 39 種類と，どの物品にも対応しないことを表す「棚」の合計 40 種類が存在する.

Pick task 用の認識では，学習データは棚に物品が 1 つ配置された画像 153 枚と，3 つ配置された画像 65 枚を用いた．物品が 1 つ配置された画像は，1 物品あたり 3,4 枚用意した．物品が 3 つ配置された画像では，物品が手前の物品によって 50%以上遮られることがないように配置しデータを集めた.

データセットを学習用と評価用に 8 対 2 で分割を行い評価を行なった所，各ピクセルに割り当てられたラベルが真値であった割合であるピクセル精度は ADAM を使わない場合が 0.866, 使う場合が 0.957 と性能の向上を確認した.

また，行為検証部の認識に利用する学習データはテーブルに置いた物品を 3D Bounding box 抽出しその上面を吸引し持ち

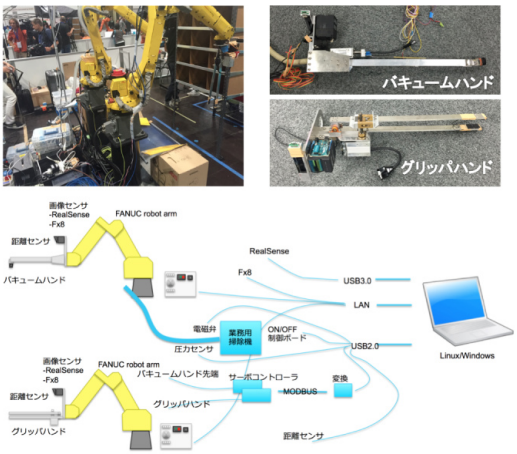

図 10　PFN のロボットシステム全体像.

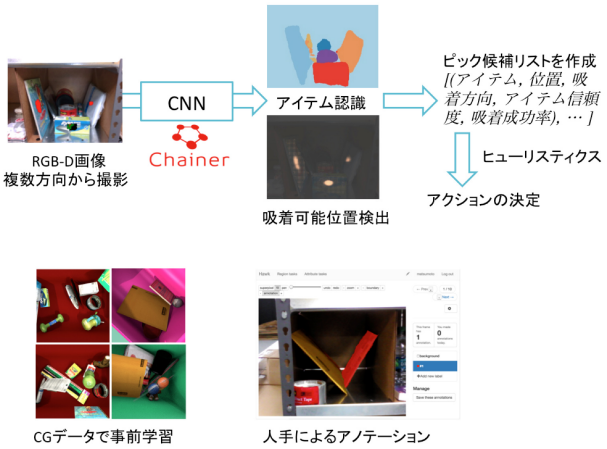

図 11　PFN のビジョンシステムの概要.

上げ，手先の部分の画像を抽出し学習データとすることで，1 物品が写っている画像を約 1000 枚収集した．またオンラインの認識では RGB-D 画像の距離情報を用いて手先の領域に存在するピクセルに対応する点以外をマスクした RGB 画像を生成し，これに対して VGG net を適用し画像中に写っている物品のラベルを識別する．

5.3　特　　徴

Team K の特徴は失敗の検出と復帰動作並びに，そのための行為検証部にある．特に後者はマニピュレーションにより物体を認識しやすい状態，すなわち，棚や箱の中の雑多な環境で隠れが存在している物体の識別を行うのではなく，それを何らかの方法で取り出し，目の前に物体を持ってくることで，物品自体を画像的にセグメンテーションしやすい状況に置き，そこで物体の識別処理を行うというものであり，画像に写っている物体を理解する受動的な Computer Vision のアプローチではなく，ロボットのアクチュエータを積極的に活用し，物体が理解しやすい状況に自身や環境を操作し，その上で対象を認識するという能動的な Robot Vision/Active Vision のアプローチを模索している点に特徴がある．また，この方式では前段の棚や箱の中の物体の認識の把持に万が一失敗した場合，あるいは，高い確信度で物体の認識を行えなかった場合でも，タスクを前に進め，後段の行為検証部で検証し認識や動作のやり直しをするという点で，継続的に稼働するシステムの構成法となっており，これは，例えば工場等での実応用において認識や把持に失敗した瞬間にラインが止まるような方法とは異なる構成であり，将来期待される複雑なタスクをおこなう作業ロボットに必須の能力と考えている．

6.　チーム紹介：PFN

株式会社 Preferred Networks の Team PFN は社員 10 名程度のチームで，ソフトウェアエンジニア，ビジョンの専門家，ロボコン経験者など多様なバックグラウンドを持つメンバーで構成されていた．3 ヶ月程度の短い準備期間であったが，Stow task で 4 位，Pick task で 2 位の成績を収めることができた．

6.1　システム構成

PFN のシステムは図 10 に示すように 2 台の産業用 6 軸多関節アーム (FANUC M-10iA) からなる．片方のアームには先端の向きを変える機構を持つ吸引式のエンドエフェクタ，もう片方のアームには挟持式のエンドエフェクタを搭載し，幅広いアイテムの種類に対応した．センサ系については，ビジョンセンサとして RGB-D カメラ (Intel Realsense SR300)，距離画像センサ (日本信号 FX8) を各アームに装備し，さらに吸着確認用に気圧センサ，棚位置のキャリブレーション用に距離センサを用いた．

ソフトウェアは各コンポーネントが ROS を介して通信する形で構成した．認識から行動を決定する部分は，Deep Learning を用いてビジョンセンサ入力からピッキング候補を計算する部分と，候補の中から最適な行動を選択するヒューリスティクスからなる．なお，Pick task と Stow task ではほぼ同じ手法を用いており，以下では両者を特に区別すること無く説明する．

6.2　ビジョン戦略

まず，吸着ハンドで取ることの出来ないアイテムであるペン立てとダンベルについては特別に場合分けを行い，挟持ハンドにより撮影→ PCL (Point Cloud Library) により位置姿勢推定→挟持トライ　というフローでピッキングを行った．

その他のアイテムについては吸着ハンドのみを用い，図 11 に示す流れで処理を行った．Realsense SR300 で撮影された RGB-D 画像を CNN (Convolutional Neural Network) に入力し，アイテム認識 (Semantic segmentation) と吸着可能位置の推定を行う．吸着可能位置とは，「そのピクセルに，大体法線方向から吸着パッドを当てた時に吸着可能かどうか」を示すマップである．なお，吸着パッドのアプローチ方向は 90 度刻みに限定し，モーションプランニングを簡略化している．

これらの結果をもとに，ある閾値以上高い信頼度で取れると判断した座標をピック候補リストに追加する．さらに複数方向から撮影した結果を統合して精度を高める (壁に立てかけてある薄い本などを取るためには斜めから撮影したデータが必要である)．最後にヒューリスティクスを用いてリストから候補を選択し (あるいは諦め)，アクションを決定する．

CNN は Fully Convolutional Encoder-Decoder Network [11] を参考にデザインした，Encoder モデルと Decoder モデルからなる 20 層のニューラルネットワークで，Encoder は共通，Decoder はアイテム認識と吸着可能位置推定で異なるモデルを用いた．入力は RGBD の 4 チャネルで解像度が 320×240，出力は，アイテム認識は 40 クラス分類 (アイテム 39 種類+背景) の解像度 320×240 のマップ，吸着可能位置推定は吸着可能性を表す 1 チャネルの 320×240 のマップである．

CNN は Deep Learning フレームワーク Chainer [10] を用いて実装した．100000 枚の CG を用いて生成した人工データを元に事前学習し，人手でアノテーションを付けた約 1500 枚の実写データを元にファインチューンを行った．ただし，CG データを用いたことによる効果は，使わない場合と比較して精度の有意な差が確認できなかった．学習には NVIDIA Titan X を用いて合計 2 日程度かかった．データセット用の写真撮影と，人手によるアノテーション作業は内製のツールを用いて 1 週間程度かけて行った．

6.3 特　　徴

PFN では RGB-D 画像からアイテム認識，吸着可能位置検出を一つのニューラルネットワークで一貫して行い，それ以降の処理は比較的単純なルールベースのヒューリスティクスで済ませた (挟持ハンドで対応した 2 アイテムを除く)．アイテムのバリエーションの多さや，複雑なオクルージョンに対応するために Deep Learning を使ったチームは多かったが，アクション候補の算出まで一気にニューラルネットで求めるアプローチを用いたのは PFN の特徴である．このようなアプローチによって，点群マッチングや把持プランニングといった複雑で調整の難しい手法を介す必要がなくなり，短期間で効率的にシステムを構築することが出来た．また，ニューラルネットワークの学習が完了した後はヒューリスティクスの調整に集中することができ，アイテムを取る優先順位や，失敗時のリトライ動作の最適化などは大きく得点に寄与することとなった．

以上のように，認識からアクションに近いところまでをニューラルネットに任せるアプローチは，精度だけでなく，開発効率の面でも大きな恩恵があった．このような手法が可能となったのは，Deep Learning の発展により，大型のニューラルネットを高精度に安定して学習させられるようになったことが大きく，今後もさらに発展が続いていくことが期待される．

7. APC の課題と今後

APC 2016 の問題設定は，第 1 回の APC2015 と比べて難しくなったとはいえ，アイテム間にオクルージョンが少ないなど，あくまでもチャレンジのための問題設定であると言える．実際の問題はさらに複雑であるため，今後はより実問題に近づけていくために，Pick task と Stow task ともにさらに難易度が高くなることが想定される．また，競技中にロボットハンドが棚に衝突して破損するアクシデントが発生するなど，ロボットの安全性についても注力する必要性があると感じた．より柔軟なピッキングやストーイングが今後の重要な課題となるだろう．Amazon Picking Challenge は，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にある．実用化に向けては様々な課題が残るが，それをオープンに共有することで，本質的な課題解決に進むことを目指している．Amazon Picking Challenge を通じて，ピッキングロボットの大いなる進化を期待したい．

文　　献

[1] N. Correll, K.E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J.M. Romano, and P.R. Wurman, “Lessons from the amazon picking challenge,” arXiv preprint arXiv:1601.05484, 2016.

[2] T. Hasegawa, Y. Yamauchi, M. Ambai, Y. Yoshida, and H. Fujiyoshi, “Keypoint Detection by Cascaded FAST,” IEEE International Conference on Image Processing, pp.5611–5615, 2014.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” IEEE International Conference on Computer Vision, pp.2564–2571, 2011.

[4] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai, “Fast Graspability Evaluation on Single Depth Maps for Bin Picking with General Grippers,” IEEE International Conference on Robotics and Automation, pp.1997–2004, 2014.

[5] S. Akizuki and M. Hashimoto, “Physical Reasoning for 3D Object Recognition using Global Hypothesis Verification,” European Conference on Computer Vision Workshops (2nd International Workshop on Recovering 6D Object Pose), vol.9915, pp.595–605, 2016.

[6] 二井谷勇佑，和田健太郎，長谷川峻，北川晋吾，板東正祐，岡田慧，稲葉雅幸，“物品棚のピッキング作業のための深層学習でのセマンティック画像分割と三次元物品輪郭抽出,” 第 34 回日本ロボット学会学術講演会講演論文集，pp.2G2–03，sep 2016.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3431–3440, 2015.

[8] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2015.

[9] K. Simoyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2015.

[10] S. Tokui, K. Oono, S. Hido, and J. Clayton, “Chainer: a next-generation open source framework for deep learning,” 2015.

[11] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang, “Object contour detection with a fully convolutional encoder-decoder network,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.193–202, 2016.