

画像認識を用いた商品棚解析 ソリューション —画像から商品の陳列情報を一括把握—

サービスイノベーション部 あかつか赤塚 はやと†隼 なかむら中村 いっせい一成
関西支社 法人営業部 こう高 そんみよん聖明

消費財業界にとって、売上の要因分析をするために店舗陳列（棚割）の実態を把握することは重要である。これまでは手作業で棚割をデータ化していたが、作業コストがかかり、また作業者にとって大きな負担となっていた。そこでドコモは商品棚画像認識技術を開発しこれを活用することで、スマートフォンのカメラなどで撮影された商品棚の写真から自動的に棚割データを作成することを実現した。これにより既存業務の作業時間の大幅な削減が可能となった。なお、本技術を活用したソリューションは、一般社団法人日本自動認識システム協会より2017年度第19回自動認識システム大賞にて優秀賞を受賞し、2018年4月よりドコモからパートナー企業への提供を開始している。

1. まえがき

消費財業界にとって、売上の要因分析をするために商品の店頭陳列（棚割）*1の実態を把握することは重要である。現在、消費財メーカーはラウンダー*2のメモ書きや、現場写真を見てドキュメントを書き起こすなどの手作業で棚割をデータ化している。しかし近年、大規模小売店の増加に伴いラウンダーの訪問店舗数は増える一方、労働人口減少により将来

的にはこれら業務の担い手が不足するといったジレンマが予想され、作業コストの削減および業務効率化が重要な課題となっている。日本全国には数万人のラウンダーがいるといわれており、これら課題の解消の観点から棚割データ作成の自動化には市場ニーズがあると考えられる。

そこで画像認識技術を応用し、商品棚を撮影した写真から自動的に棚割データを抽出することが提案されている。しかしながら既存の画像認識技術には、

©2018 NTT DOCOMO, INC.
本誌掲載記事の無断転載を禁じます。

† 現在、イノベーション統括部

*1 店頭陳列（棚割）：商品棚にてどの商品がどの位置に陳列されているかのレイアウトを指す。

*2 ラウンダー：スーパーやデパートなどの店舗を巡回し、自社商品の販売促進活動を実施する担当者を指す。具体的な業務として店舗にて自社製品の販売・提案・展示状況の把握、他社製品の調査、販売店からの要望のヒアリングなどを実施する。

実店舗での利用において主に2つの課題があった。

- ・1つめは圧縮陳列*3された商品の物体検出が難しいことである。日本のように陳列スペースが限られている場合、シャンプーなどの詰替え袋は圧縮陳列されているため形状が歪み、見え方が一定ではない。
- ・2つめは商品が必ずしも正面を向いていないため、商品の認識が難しいことである。実店舗では同じ商品であっても、さまざまな方向を向いて陳列されている。

従来のソリューションにおいては、店頭で商品の被りをできるだけなくし、パッケージを正面に向けて陳列させるなど撮影方法を工夫することで、画像認識がしやすい環境を整えていたが、このような対応は時間がかかるため実用的ではなかった。

そこで今回ドコモは、圧縮陳列や撮影角度の課題を解決し、商品棚を撮影した写真から商品の棚割データを自動的に解析するディープラーニング技術を利用した商品棚画像認識エンジン（以下、本画像認識エンジン）を開発した。本画像認識エンジンは、一般社団法人日本自動認識システム協会より「商品棚を画像認識から再現する新規性」と「ソリューションの実現性」の観点が評価され、2017年度第19回自動認識システム大賞にて優秀賞 [1] を受賞し、2018年4月よりファーストユーザとして株式会社サイバーリンクスにその提供を開始した [2]。本稿では、本画像認識エンジンの技術内容と具体的な利用シーンについて解説する。

2. 画像認識概要

本画像認識エンジンは2つの技術から構成される。

1つめはディープラーニングを用いた物体検出技術であり、写真上のどこに棚板や商品があるのかを検出することで、商品の空間配置を解析することが

できる。図1に表示された枠が物体検出の結果である。ドコモの物体検出技術はディープラーニングを用いてさまざまな実店舗の陳列状態の画像を学習することで、商品が狭いスペースに圧縮されて陳列されている場合であっても商品を高い精度で検出できることが特長である。

2つめは局所特徴量*4を用いた特定物体認識技術であり、前述のように検出された商品領域の部分画像と、事前に画像データベースに登録された大量の商品画像データとを比較することによって、棚上の各物体領域がどの商品であるかを識別することが可能である。ドコモの特定物体認識技術は、さまざまな角度の画像を大規模な画像データベースにあらかじめ登録しておくことで、正面以外の画像であっても高速にかつ高い精度で認識ができることが特長である。

これら2つの技術による処理フローを図1に示す。

2.1 物体検出技術

(1) アルゴリズム詳細

ドコモが開発した物体検出技術は、ディープラーニングを応用し、画像中の各物体の位置（矩形の左上と右下の座標）と種類を推定することができる。そのため、事前にどのような物体を検出したいかを物体検出エンジンに機械学習*5させておく必要がある。機械学習には画像データとアノテーションデータ*6を数百～数千枚単位で用意した。アノテーションデータの内容は、図2に示すように、画像中の各物体がどの領域に描画されているかを表す位置情報（矩形の左上の座標（ x_{min} , y_{min} ）と右下の座標（ x_{max} , y_{max} ）と物体の種類情報（例えば、柔軟剤、洗剤、飲料水、棚板など）である。

機械学習の結果、学習済みモデルが作成され、それを物体検出エンジンにて読み込むことで「推論*7」処理を行うことが可能となり、入力された未知の画像中の物体を検出することができるようになる。

*3 圧縮陳列：狭い売り場空間を商品で徹底的に満たし陳列する方法であり、商品が多少潰れて陳列されることが多い。

*4 局所特徴量：データから抽出される、そのデータの特徴づける量（数値）のこと。本稿における特徴量とは特に画像特徴量とも呼ばれ、画像から検出された特徴点（コーナー点）においてその周辺の輝度分布を特徴づける量である。

*5 機械学習：人間が、知覚、経験から知識や判断基準、動作などを獲得していくように、コンピュータにデータから知識や判断基準、動作などを獲得させる技術。

*6 アノテーションデータ：ここでは画像データが何かを示すメタデータを指す。

推論処理は、主に次の4つのフェーズに分かれている (図3)。

①特徴量の抽出

ユーザから入力された画像に対して、畳込



図1 商品棚画像認識処理フロー



図2 学習に必要な画像とアノテーションデータ例

*7 推論：ここでは事前に学習したモデルを利用し、入力された画像が何かを推定する処理を指す。

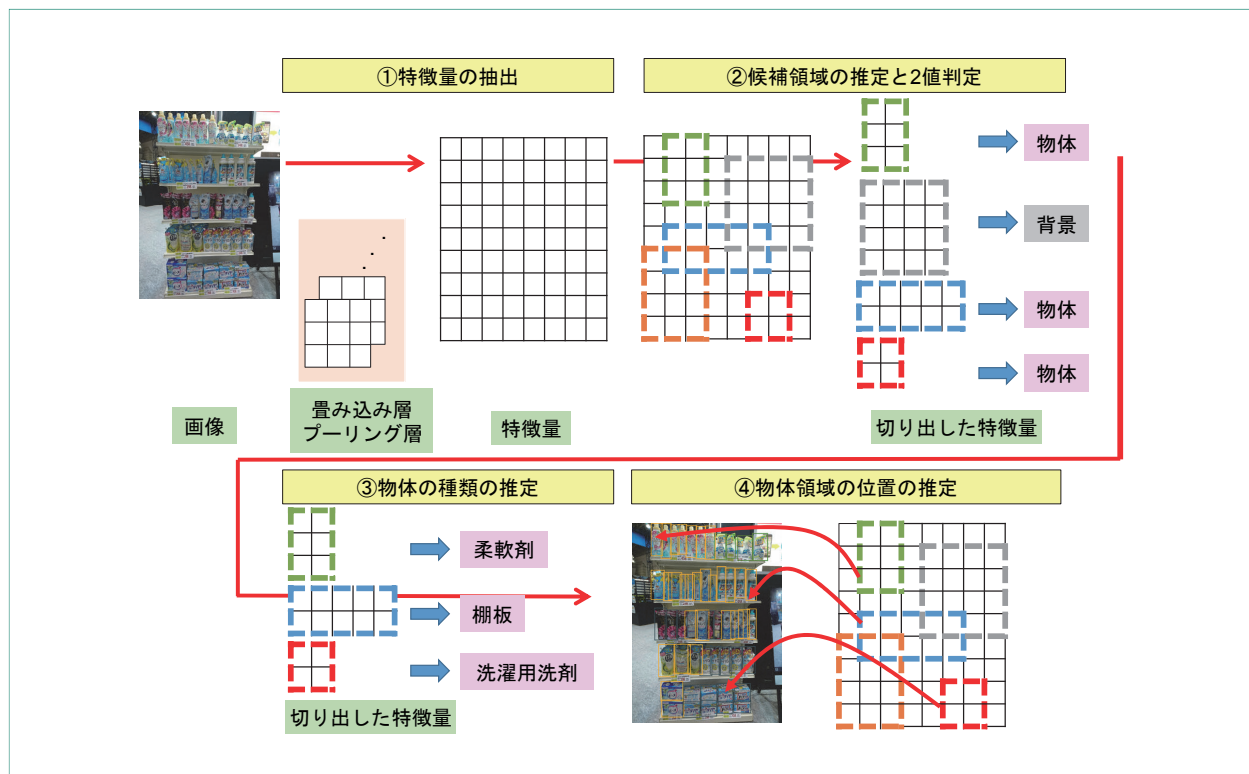


図3 検出アルゴリズムの処理の流れ

み*⁸やプーリング*⁹を複数回行い、特徴量化する。

②候補領域の推定と2値判定

上記で抽出した特徴量を用いて、さまざまなアスペクト比・スケールで複数の領域を切り出し、各領域に対して全結合層*¹⁰を適用し、「物体」であるか「背景」であるかの2値判定を行う。

③種類の推定

2値判定の結果(②)で「物体」と推定された領域に対して、②とは重みが異なる複数の全結合層を適用し、物体の種類を推定する。このとき複数の種類の候補とその確率が得られるので、最も大きい確率のものをその領域の種類とする。

④画像内での物体領域の位置の推定

前工程(③)で求めた物体の種類と、

画像上の位置、特徴量を用いて、画像内での物体領域の位置情報を推定する。

以上により、1枚の画像中に写る多数の物体の位置と種類を推定して出力する。

なお、本手法は、商品だけでなく棚板も検出可能である。これにより、棚の段数や各商品が何段めのどの位置に存在するか、また、積み上げられている同一商品の数を含めて、推定を行うことができる。

(2)認識精度

商品が多数配置してある商品棚の画像を数千枚用いて機械学習させ、学習に用いていない実店舗で撮影した画像100枚で物体検出エンジンに推論をさせ、物体検出技術の精度評価を行った。商品は、洗剤、柔軟剤などの洗濯用品と、ペットボトル飲料水や缶ビールなどの飲料水の2つのカテゴリーを用いた。精

*8 畳込み：画像や特徴量などの入力に対して横方向と縦方向に走査しながら特定のサイズのベクトルを掛け合わせた値を出力すること。掛け合わせたベクトルと類似するパターンを抽出する。

*9 プーリング：特徴量に対して横方向と縦方向に走査しながら特定のサイズ(2×2や3×3など)内の最大値や平均値の値を出力

すること。特徴量の次元を削減し、推論時のロバスト性を高める。

*10 全結合層：特徴量のすべての数値に対して重みを掛け合わせ1つの数値を出力する際の重みのこと。畳込みで抽出された特徴量に対してこの重みを掛け合わせることで特徴量を変換し、商品のカテゴリーを分類する。

度指標には、Recall（正しく検出した商品数／全商品数）とPrecision（正しく検出した商品数／検出した商品の数）を用いた。精度評価の際にはまず、推定した各領域がどのGround-truth領域（正解領域）を推定したかを判断する必要がある。そのために図4に示す、2つの領域のオーバーラップの程度を0～1.0の数値で表す指標であるIoU overlapを用いる。このIoU overlapは値が大きいほど2つの領域のオーバーラップの程度が大きいことを示す。このIoU

overlapが0.6以上となり、かつ推定した物体の種類とGround-truthの物体の種類が一致した場合のみ正解とした。ただし、③で求めた物体の種類が0.6未満である領域は確信度が低いと考え除去した。

この条件で評価を行ったところ、飲料水の検出精度はRecallが91.2%、Precisionが92.1%となった。洗濯用品の検出精度はRecallが92.0%、Precisionが99.7%となった。図5に示すように実店舗に置いてある商品はパッケージが正面を向いているとは限ら

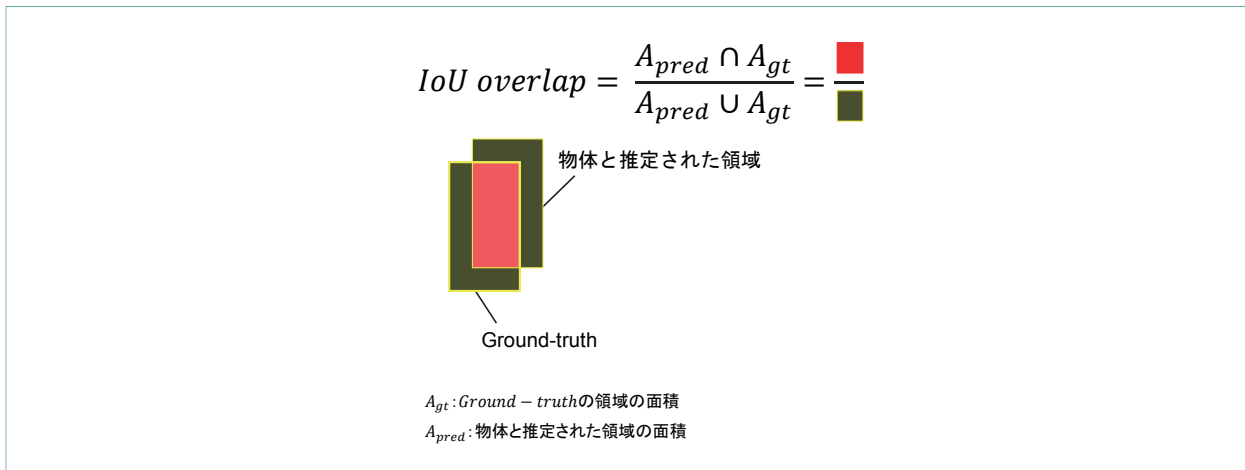


図4 IoU overlap

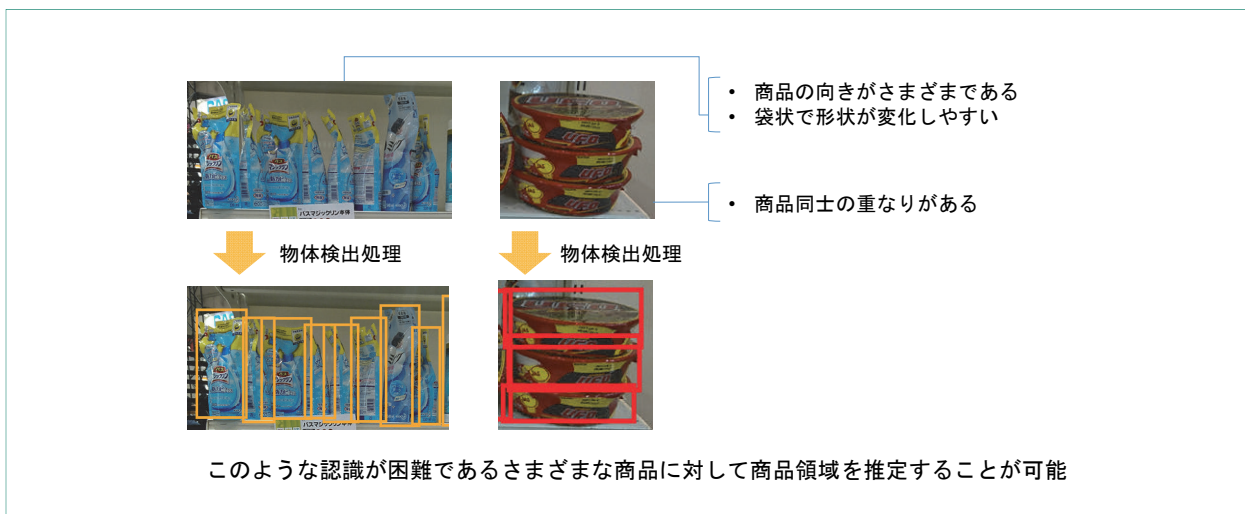


図5 商品棚の商品の特徴

ず、また、袋状のパッケージのような形が変化しやすい商品が存在し、さらに商品同士が重なることで一部隠れる場合があるが、本技術は高精度で検出が可能である。

(3)処理速度

検出にあたって、GPU、CPUそれぞれを用いた場合の処理速度を計測した。NVIDIA Tesla^{*11} M40のGPU1基を用いて推論を行ったところ約0.3秒で処理ができた。一方、GPUを利用せずCPU (Intel[®] Xeon[®]*12 CPU E5-2630L v3 @1.80GHz) 上ですべての演算処理を行った場合は、約7.0秒の処理時間がかかり、GPUを利用することで10倍以上高速化が可能であることを確認できた。利用シーンによってシステム要件は異なるが、現状ではリアルタイムな処理にGPUの利用が必須であると思われる。なお、ここで記載の処理速度は入力する画像ファイルサイズに依存するため、画像サイズが小さい場合はより短時間で処理が可能である。

2.2 特定物体認識技術

(1)アルゴリズム詳細

特定物体認識技術は、入力された画像と、画像

データベースに事前に登録されている画像とを比較し、入力画像と類似する登録画像を判別することで、入力画像が何であるかを識別する技術である (図6)。図1の商品棚画像認識において、物体検出された枠内の部分画像が特定物体認識への入力画像となる。事前に複数の角度から撮影された商品画像をデータベースに登録することで、画像認識エンジンはさまざまな角度で陳列されている商品の認識が可能になる。ドコモの特定物体認識のアルゴリズムの詳細については文献 [3] を参照されたい。

(2)システム要件

ドコモの特定物体認識技術は、数百万個の画像が画像データベースに登録されていてリアルタイムに処理が可能であるが、そのために画像データベースをメモリ上に展開する必要がある。どの程度メモリが必要かは、登録する画像枚数と1枚当りに割り当てる情報量の設定値に依存するが、物理サーバにはメモリの上限があるため、1台のサーバで保持できる画像数には上限がある。そこでサーバを複数台立ててスケールアウト^{*13}させることで、大規模な画像データベースを構築することができる。特定物体認識はCPU上で高速に動作が可能であり、エン

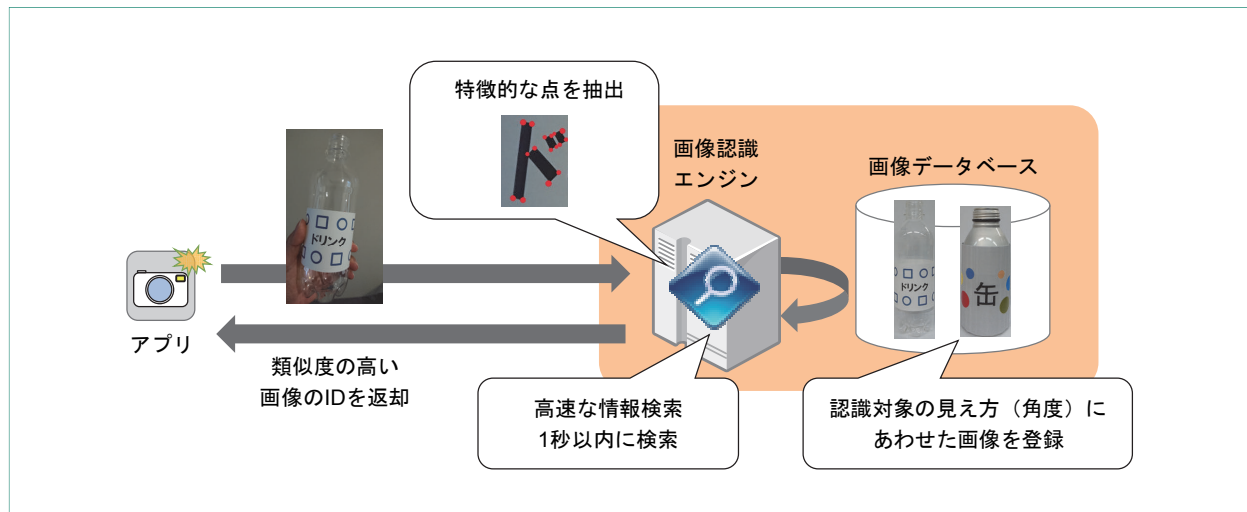


図6 特定物体認識

*11 NVIDIA Tesla：米国および/または他国のNVIDIA Corporationの商標および/または登録商標。

*12 Intel[®] Xeon[®]：アメリカ合衆国および/またはその他の国におけるIntel Corporationまたはその子会社の商標。

*13 スケールアウト：サービス要求の増加などにより、網内の処理量が不足した際に、新たなリソースを追加して割り付け、処理量の増強を行う。

ジンの設定値および登録画像枚数にも依存するが、1枚当たり数百ミリ秒で認識が可能である。

(3)認識精度

実験環境に商品棚を準備し、実店舗を再現する形で商品を陳列することで認識精度を評価した。商品の種別としては、麺（カップ麺、袋麺）、洗剤（容器、詰替え用袋）、飲料水（缶、ペットボトル）を利用した。数百枚の商品画像を用いて評価した結果、第1位の候補が正解である確率は95.96%となり、高い精度を確認できた。

一方、認識が失敗したものとして、サイズ違いの商品や、色違いの商品、撮影の影響で手振れしている画像があった。サイズ違いや色違い、文字が一文字だけ異なっている商品などは、人間が見ても間違えるほど見た目が酷似しているため、一般に画像認識による識別は困難である。手振れに関しては、画像から必要な情報の抽出ができなため認識精度が

落ちる。上記以外に、画像の解像度が低く、商品領域の画像が小さくなった場合には、抽出可能な画像情報量が少なくなるため、認識精度が落ちる傾向にある。認識精度をあげるためには、手振れを極力なくし、できるだけ商品棚に寄って高精細なカメラで撮影する必要がある。

3. 棚割データ作成

本画像認識エンジンを利用することで、商品棚の写真から商品の特定および商品の陳列状況を解析し、その結果を棚割データとして出力することができる。この棚割データを棚割のシミュレーション作業を行うソフトウェア（棚割ソフト）へ読み込ませると、棚割の可視化・分析が可能である（図7）。なお、日本国内の代表的な棚割ソフトとして、株式会社サイバーリンクスが提供する棚POWER[®]*14 [4]、日本総合



図7 棚割りデータの可視化

*14 棚POWER[®]: (株)サイバーリンクスの登録商標。

システム株式会社が提供するStoreManager[®]*15 [5]がある。

日本国内では、棚割データを表現する共通フォーマットとして「共通棚割情報（PTS：Planogram Transfer Specifications）」[6]がある。PTSでは、商品ごとにフェイス数*16、積上げ数、どの段に位置するか、などを可視化するために必要な情報を含んでいる。前述の物体検出技術は商品以外に棚板の検出も可能なので、棚板（棚段）ごとに商品のフェイス数、積上げ数を正確にカウントして、PTSの仕様準拠した棚割データを作成することができる。これにより、棚割ソフトに読み込んで棚割を可視化するとき、あらためて棚段の設定をしなくても、元の商品棚を正確に再現することができる。

4. ユースケース

前述の通り、本画像認識エンジンはラウンダーの棚割りデータ作成に適用可能である。我々の見積りでは、写真から棚割データを自動的に作成することによって、手作業による棚割データ作成方法と比較し、作業時間を10分の1に短縮可能と考えている。さらに、作業コストの削減のほかに重労働を緩和する効果もある。具体的には、現状の棚割りデータ作成時に行うバーコードリーダーの読取り作業は、商品棚の上の段から下の段へと上下の動きが非常に多く、足腰に非常に負担がかかる。一方、本画像認識エンジンを用いれば、正面から2、3枚写真を撮るだけで済むので、作業を簡略化し負担の低減が図れる。

また、上記以外の利用シーンとして、消費財業界のラウンダーや、小売りの店舗スタッフが作成する業務報告書への写真の利活用がある。現状では、実店舗において商品の陳列業務など作業が完了したことを報告するために商品棚を撮影しているが、写真は目視確認レベルの簡易な利用に留まっている。商品棚の写真から自動的に棚割データの作成ができれば、

ば、本部やリモートオフィスにおいて、離れた実店舗の陳列を調べ、誤りがないかを確認したり、店舗独自の陳列の工夫を吸い上げたりするような管理業務の支援も可能である。

5. あとがき

本稿では、本画像認識エンジンについて、商品と棚板を検出する物体検出と、商品特定する特定物体認識の2つの要素技術を解説し、またそれを活用して、商品棚画像から商品の特定、商品の陳列位置、棚板の検出を実施することで、棚割データの自動作成が可能であることを解説した。

本画像認識エンジンの課題の1つは、実店舗での技術検証および精度改善である。具体的には前述の認識が難しいサイズ違い、色違い、文字などの部分違いの商品にも頑健な特定物体認識の技術検討を進める必要がある。また実店舗では、棚板にプライスレール*17が付けられており、商品の一部が隠れて情報が欠損するため、認識精度が劣化する場合も存在する。今後は、より多くのパートナー企業と連携して技術検証を進め、そのような認識が困難なケースを洗い出して認識アルゴリズムの改善を行っていく。

またドコモでは、商品棚画像認識以外にもスポーツ映像解析 [7]、ARサービスへの適用 [8]、名刺のデータ化に伴う業務効率化への適用 [9]、ドローンを用いた海岸保安林の松枯れ検出 [10] など、幅広い分野で画像認識の、自治体・パートナー企業への提供を行っている。これらに関して条件はあるものの、アプリ開発／サービス開発に利用できるAPI（Application Program Interface）*18として、ドコモが開発した画像認識をdocomo Developer support [11] より一部公開しており、会員登録および利用申請を行うことで利用が可能である。今後も、ドコモは商品棚画像認識を含めさまざまな分野にお

*15 StoreManager[®]：日本総合システム(株)の登録商標。

*16 フェイス数：流通用語で商品棚を正面から見たときに陳列している商品の数をフェイス数という。

*17 プライスレール：商品棚の棚板ごとに設置されている値札が張られているレール。

*18 API：ソフトウェアの機能を他のプログラムから利用できるように切り出したインタフェース。

いてパートナー企業への価値の提供を目的に画像認識の技術開発を行っていく。

文 献

- [1] 一般社団法人日本自動認識システム協会：“自動認識システム大賞 | JAISA 一般社団法人 日本自動認識システム協会.”
<http://www.jaisa.jp/award.php>
- [2] NTTドコモ報道発表資料：“（お知らせ）商品棚画像認識エンジンを開発—ドコモのAI技術により、商品棚の配置状況が一目でわかる—,” Mar. 2018.
https://www.nttdocomo.co.jp/info/news_release/2018/03/01_01.html
- [3] 赤塚, ほか：“高速大規模画像認識エンジンの開発とAPIの提供,” 本誌, Vol.23, No.1, pp.14-20, Apr. 2015.
- [4] 株式会社サイバーリンクス：“基本操作 | 棚POWER | 株式会社サイバーリンクス棚POWER・店POWER.”
https://www.tanapower.com/tpower/basic_operation.htm
- [5] 日本総合システム株式会社：“StoreManager | 日本総合システム株式会社.”
<https://tanawari.jp>
- [6] 棚割システム研究推進会：“共通棚割情報（PTS）のご紹介：棚割システム研究推進会.”
<https://www.planet-van.co.jp/planogram/pts/index.html>
- [7] サッカードットコム株式会社, NTTドコモ：“AI技術「スポーツ動画センシング」を開発, マルチアングル自動録画システムと連携したトライアルを開始～自らのプレーを手軽に映像で振り返る新たなスポーツ体験を提供～,” Feb. 2018.
https://www.nttdocomo.co.jp/binary/pdf/info/news_release/topics/topics_180201_00.pdf
- [8] NTTドコモ：“サイバネットシステム株式会社「cybARnet」へ「画像認識システム」を提供,” Oct. 2015.
https://www.nttdocomo.co.jp/binary/pdf/corporate/technology/rd/topics/2015/topics_151001.pdf
- [9] NTTドコモ報道発表資料：“（お知らせ）Sansan社がドコモの画像認識システムを採用,” May 2016.
https://www.nttdocomo.co.jp/info/news_release/notice/2016/05/25_00.html
- [10] NTTドコモ報道発表資料：“新潟市におけるドローン実証プロジェクトに関する連携協定を締結,” Sep. 2016.
https://www.nttdocomo.co.jp/info/news_release/2016/09/21_00.html
- [11] NTTドコモ：“画像認識 | docomo Developer support | NTTドコモ.”
https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_docs_id=102