360度動画中における商品認識

藤原　夏姫†　　横山　昌平†

† 東京都立大学院システムデザイン研究科情報科学域　〒 191–0061 東京都日野市旭が丘 6-6
E-mail: †natsuki-fujiwara@ed.tmu.ac.jp, ††shohei@tmu.ac.jp

あらまし　近年、一度に全方位の撮影が可能となる 360 度カメラが普及進んでいる。360 度カメラは、VR(Virtual Reality:仮想現実) や AR(Augmented Reality:拡張現実) など様々なシーンの仮想化を可能にするためにも使用されている。また、企業が購入者の目を引くコンテンツとして商品紹介の動画などにも 360 度カメラは用いられている。しかし、これらの動画内では動画中に現れる商品の種類やブランドに関する表示がないため視聴者は商品に関しての情報を文字媒体により動画から得ることが難しい。そこで本研究では、360 度動画中の各商品と商品のロゴの検出を YOLOv3 を用いて行おうと試みた。ここで、YOLOv3 とは機械学習を用いた既存の物体検出手法である。提案手法の学習に用いるアルゴリズムは既存の手法を利用する。また、今回認識させる商品はカバンのみであり、カバンと様々なブランドのロゴを学習させることでカバンとロゴの位置を検出する。

キーワード　マルチメディア, 機会学習, 可視化,360 度カメラ, 物体認識

1　はじめに

近年コロナ感染症の影響で活動自粛を強いられる場面が多々あり、安心して旅行や買い物に行けない現状がある。そのため、動画を視聴することで海外旅行をしている気分になったり EC サイトなどを利用して商品を購入したりする機会が増えている。例えば楽天では、2020 年の第一四半期・第二四半期と共に楽天市場をはじめとするショッピング E コマースでの売上が全年比より 50 %程度上回っている [1]。また、独自のオンラインショッピングシステムを持つニトリホールディングス社では、コロナ感染症により外出自粛が設けられていた時期も含まれる 2020 年 3 月-8 月の通信事業売上高は、前年同時期の 56.4 %増しとなっている [2]。これらの企業を始めとして多くの EC サイトでは、商品を紹介する際に画像を使用している。また、使用している画像は、商品を複数の角度で撮影したものを複数枚のみ用いているため全方位から商品を確認出来ない点に加えて、お店の雰囲気を味わえない。この課題を解決出来る方法の 1 つとして 360 度カメラによる仮想店舗を体験出来るシステム開発が考えられる。

手軽に全方位撮影できる 360 度カメラの普及が近年進んでいる。RICOH 社の THEAT シリーズ [1]や Samsung 社の Gear 360[2] 、GoPro[3]などの 360 度カメラが発売しており、誰でも簡単に全方位画像や動画が撮影出来るようになった。360 度カメラとは、複数のレンズを使い撮影して画像を繋ぎ合わせることで撮影者の周囲の空間の撮影や商品を 360 度自由に視聴できることを可能にした。また、360 度カメラで撮影された動画像は、正距円筒図法に変換されて全天球パノラマ動画像を作成して保存または配信される。正距円筒図法とは、緯線と経線が直角で等間隔に書かれている世界地図のように変換する方法である。全天球パノラマ動画像とは平面画像であり、通常の動画像と同じ処理が可能である。そして、360 度カメラを用いることで周囲のシーンを全て撮影できるため、臨場感や没入感を味わうことが可能である。利用例としては、Google Street View[4] や、不動産の内見での使用、企業の広告コンテンツなどが挙げられる。また、360 度カメラにより撮影された動画を利用することで、AR(Augmented Reality:拡張現実) や (VRVirtual Reality: 仮想現実) のような仮想的な体験も実現させることが可能である。　そして近年、360 度動画を YouTube[5]や Instagram[6]などの SNS にアップすることが非常に簡単になっている。観光地や店内を体験できる動画や、アトラクションを体験できる動画など多様なジャンルで使用されている。360 度動画を利用することで、実際に店舗にいないのにも関わらずに店舗にいるような感覚でお店の雰囲気を楽しみながら商品を視聴できるような仮想店舗の実現が可能であると考えている。しかし、単純に 360 度カメラで撮影しただけでは動画内において文字による商品説明や商品のブランド名などの記載がないため、商品に関する情報が動画視聴だけでは得られない。また、実際に店舗で買い物をしている状態に近づけるためには、商品のロゴに関する情報も必要だと考えた。実店舗で買い物をする際には、商品のロゴやマークなどから商品を探す経験は多くの人が経験しているためだ。また、企業にとってロゴは商品の認知度を向上させたりグローバル企業においてはローカル戦略の効果も期待できたりと商品を売るためには非常に重要なものである。[3]

そこで我々は、360 度動画を正距円筒図法で変換した動画における物体検出に加えてロゴの検出も試みた。物体検出とは、入力した動画や画像に対して、物体の位置と種類を検出する技術

1：https://theta360.com/ja/
2：https://www.galaxymobile.jp/gear-360/
3：https://gopro.com/ja/jp/
4：https://www.google.co.jp/maps/preview
5：https://www.youtube.com/
6：https://www.instagram.com/

である。近年物体検出を行う手法として、Region-based Convolutional Neural Network(以下 R-CNN)、Yolo Only Look Once version3(以下 YOLOv3)、Single Shot Multibox Detector(SSD) など機械学習による物体認識を行う手法が多く提案されている。機械学習による物体検出の仕組みは、ラベル付けされあた大量のデータセットを機械学習することで、そのクラスにおける特徴量を自動で抽出し、未知の入力動画中の物体を検出できるようになっている。応用事例として、製造業における不良品の発見や無人レジなど幅広い業界・分野で用いられている技術である。この技術を用いることで、効率的に動画内における物体検出が可能となる。　本研究では、既存の物体検出方法 YOLOv3 を用いて 360 度動画内の商品認識を行う。また、今回は複数の種類が異なる物体を検出するのではなくて、カバンの検出とカバンのロゴの検出をしようと試みた。　本論文の構成は以下の通りである．2 章では，関連研究について述べる。3 章では，提案手法について述べる．4 章では，実験の結果と考察について述べる．5 章では、通常のカバン以外のロゴをもつ物体の画像に対してロゴを検出出来るのかを確かめた。6 章では，本研究のまとめと今後の課題について述べる．

2 関 連 研 究

本章では関連研究について述べる。本研究のように 360 度カメラを用いてはいないが、画像中のロゴや商品の物体検出や、商品の個数や位置を推定する研究は他にも行われている。

2.1 Selective Search と AKAZE 特徴量を用いたロゴ検出

西本 [4] らの研究では、AKAZE 特徴量 [7]を用いた特徴点マッチングによるロゴ検出を行う手法を提案した。特徴点マッチングとは、異なる画像間において固有の点を対応付けることであり、特徴点の検出を行い、特徴量を記述してマッチングを行う 3 ステップを踏む必要がある。また、固有の点座標は、角や線のカーブなど特徴的な部位を表している。ロゴ検出のために、ロゴがある特定の商品画像と特定のロゴ領域のみが映った画像を用いて AKAZE 特徴量をそれぞれ計算した。また、マッチング距離を特徴量間のユーグリッド距離の平均値として、設定した閾値以下となった場合に商品画像中にロゴが含まれるとすることでロゴ検出を試みた。ここで、商品画像においては Selective Search により物体らしい領域を分割することでロゴ領域を事前に絞ってから特徴量を算出した。この手法は、特定のロゴに対して特徴量を算出しているため、他のロゴの検出はその都度特徴量を算出する必要があり複数のロゴを検出させたいときに計算量が膨大となる。また、未知のロゴの商品に対してロゴがありそうな場所を検出するのは不可能である。

2.2 ニューラルネットワークによる物体認識技術の商品管理システムへの応用

井岡 [5] らの研究では、食品工場の生産ライン管理において食肉商品の中身と商品のパッケージのラベルが正しいのかを YES/No/不明で判定するシステムを提案した。YOLO により商品画像中においてラベルの位置を検出して、Residual Network(ResNet) を用いてラベルに書かれている内容をカテゴリー別で判別するラベル判別モデルを作成した。また、ResNet により食肉商品の内容物をカテゴリー別で判別する内容物判別モデルも作成した。両モデルにおいて、カテゴリー結果と判別スコアが出力される。判別スコアが閾値以下だった場合には不明と認識されて、閾値以上だった場合にはカテゴリー名が一致した場合には YES、しなかった場合には NO と出力させることでラベルと内容物が一致しているかを最終的には表示させるシステムを構築した。また、ラベルが新しくなったり商品トレイの色や形が変わった場合は認識精度が悪くなった。この手法は特定の商品に対する特定のラベルの場合は精度よく認識できる。しかし、多様な商品でラベルが異なるデザインの際には精度よく認識できない。

2.3 反教師あり学習による商品画像中の個数と位置の同時推定

藤橋 [6] らの研究では、CNN を用いて商品の位置や大きさの情報を必要とせずに画像中の商品の個数と位置を推定する手法を提案した。画像中に存在する商品の個数を入力して、出力は商品を矩形で囲って表示するものとなっている。CNN の学習では、ランダムに与えられた 2 点の座標により四角形を生成する。入力画像・生成した四角形の内部と外部をそれぞれ CNN により畳み込み処理を行った。また、用いた CNN のネットワークは深みが 16 層である VGG16 を用いて、最終出力のノードは 1 つとした。学習した CNN を用いて入力画像に対する物体の推定個数を求める。また、Selective　Search によって複数の物体らしき候補となる矩形ごとに対して矩形内部と矩形外部を切り取った画像を生成する。生成された画像に対して、学習した CNN を用いて矩形内外の推定個数を得る。この 3 つの推定個数と推定個数の重なり具合を泡らしている IoU を基に物体の推定位置を求めた。この手法では、商品の個数と位置のみを知りたい場合には最適だが商品の種類やどこの商品なのかを知ることが出来ない。また、Selective Search によって物体らしい領域を検出してからそれぞれに対しして計算するため、計算量が膨大となる。

2.4 深層学習を利用したウメ「露茜」の画像による熟度分類

建本 [7] らの研究では、SSD によりウメの一種である露茜の熟度分類を行う手法を提案した。物体を検出して、検出されたウメに対して CNN を用いてウメの熟度を判定する。画像中のウメの位置を矩形で表していて、さらにウメの熟度を表示するものである。教師データとして、ウメの熟度を着色がどの程度進んでいるかで 5 段階に分けてラベリングをしているデータを用いている。SSD による物体検出の精度は 98.9 %であり、CNN を用いた熟度の誤判断は 4 %であった。ここで、SSD の物体検出の仕組みは、画像上に大きや形の異なるデフォルトボックスを載せて、各ボックスに対して物体の位置の予測とクラスの予測を行い、各予測の最も高い予測値を基にして IoU を

7：http://www.robesafe.com/personal/pablo.alcantarilla/kaze.html

計算することで物体検出を行っている。そのため、スケール変化に強く、複数のオブジェクトがシーン内に存在していも精度が良い検出が可能となる。建本らの研究では、この利点を生かしてウメの検出を進めた。しかし、SSD で物体検出を行う際には基本矩形のサイズの設定や位置予測の閾値など恣意的なパラメータを設定する必要がある。

3 提 案 手 法

本研究では、全天球パノラマ動画上のカバンのロゴとカバンを検出をさせるための手法を提案する。カバンと共にカバンのロゴを学習させることで検出を行った。本研究で提案する手法の流れを図 1 に示した。また、用意したデータセットや学習、検出アルゴリズムなどに対して特別な処理は一切加えていない。通常の動画や画像での学習・検出に用いられている既存のアルゴリズムを用いることでロゴの検出行う。検証用の 360 度動画の撮影は RICOH 社の 360 度カメラ THETA を用いて行った。また、学習用のカバンの画像は、Flickr API を用いて取得した。そして、ロゴ検出においては物体の位置と種類を検出する機械学習モデルである YOLOv3 を用いて実験を行う。

図 1 提 案 手 法

3.1 学習データの収集

本研究では世界中で利用されている画像共有サイトである Flickr[8] と海外ファッション通販サイトである BUYMA[9] を用いて画像を収集する。Flickr が提供している API を用いて取得したカバンの画像に加えて、BUYMA をスクレイピングすることで取得したカバンの画像を使用した。本研究では、カバンに関する画像群を収集するためにプログラムを作成して自動的に収集した。今回はロゴを学習する必要があるため、ロゴマークがあるカバンのブランドを事前に指定した。取得したブランドは、ロゴが目立つ 8 社のブランドを中心にカバンの画像を合計で 4080 枚取得した。内訳は表 3.1 以下である。表 3.1 ロゴなしの項目があるが、ブランドのロゴが明確に見えなかったりロゴがないカバンの件数である。

3.2 YOLOv3

YOLOv3 は、入力画像から物体の座標と種類を検出する機械学習を用いた物体検出手法である。以前のモデルは、物体らしい領域候補を複数挙げて、各領域に対するクラス分類を行ってきた。しかし、YOLO は物体領域候補をあげずに、1 つの CNN で物体の検出からクラス分類までを予測するモデルになっている。また、物体の位置情報とカテゴリーを記載したデータセットを与えて学習させて、物体の座標と大きさを直接予測する回帰問題として取り扱っている。物体の存在領域は特定の領域に何らかの物体が存在するという条件のもとで、どのクラスに属するのかを確立を用いて表現している。YOLO を用いることで、処理が早く画像全体を見ているため精度が良くなる利点が挙げられる。 本研究では、Keras-yolov3 を使用した。Coco Dataset を用いて 80 クラスの学習済みモデルをダウンロードして利用した。図 2 に YOLOv3 での物体検出を行った画像を記載した。ここで用いた画像は、iPhone7 によって撮影された画像である。また、図 4 に Rico 社の Theta V を用いて撮影した全天球パノラマ画像において既存の YOLOv3 モデルにより物体検出を行った結果を示した。360 度画像においても、通常の画像と同様に物体検出が出来ていることを確認した。今回は、独自のデータセットを用いてカバンの検出に加えてロゴの検出も行う。物体検出対象は、カバンとロゴのみである。また、モデル学習の際には GPU を使用して効率化を図った。

表 1 取得データ数

ブランド名	枚数	ブランド名	枚数
MICHAEL KORS	686 枚	Louis Vuitton	471 枚
Adidas	561 枚	Prada	390 枚
Coach	661 枚	Chanel	362 枚
The North Face	97 枚	Gucci	454 枚

図 2 YOLOv3 での物体認識

図 3 全天球パノラマ画像における YOLOv3 での物体認識

8：https://www.flickr.com/groups/japanese/
9：https://www.buyma.com/

図 4　アノテーション画像

3.3　アノテーション

Flickr で取得した全てのカバンの画像に対してアノテーションを行う。画像の教師データを作成するアノテーションツールとして、本研究では Microsoft が無償提供する Vott を用いた。図 4 のようにロゴやカバンの存在領域をマウスで指定することで、クラス番号と座標データの情報が XML ファイルに保存される。YOLO を動かすためには、クラスの位置情報が記載されている XML ファイルを txt ファイルへ変換する必要がある。最終的な txt ファイルは、画像パス/左上の座標/右上の座標/クラス番号の形式となるように各画像に対応して保存する。

3.4　学　　習

画像と物体の位置情報を示す座標データが記載された txt ファイルを教師データとして深層学習をする。今回は、ロゴとカバンのクラスを作成して新たに再学習した YOLO モデルのみを用いて、カバンとロゴの認識を行った。

4　評 価 実 験

本章では、360 度カメラによって撮影された動画に対してロゴとカバンの検出をする実験を行った。RICOH THETA Z1 を用いて、約 15 秒の動画を撮影した。この際に、カバンの数を変えて 2 種類の動画を撮影した。また、今回使用した物体は、Michael Kors・Louis Vuitton・Adidas のようにロゴがはっきり確認できるカバンとロゴが確認できないカバンやスーツケースなどを使用した。

4.1　検 出 結 果

入力動画は、正距円筒図法で変換した約 15 秒 の全天球パノラマ動画である。図 5 と図 6 に今回使用した入力動画を示した。入力動画 1 では、7 個ののカバンを使用した。入力動画 2 では、8 個のカバンを用意した。入力動画 1 と 2 で使用したカバンのロゴは教師データとして用意していないものも含まれている。この 2 つの動画の検出結果の一部を図 7 と図 8 に表示した。出力結果より、ロゴとカバンは認識出来ていることが確認できた。しかし、全てのカバンやロゴに対して認識できていない。

図 5　入力動画 1

図 6　入力動画 2

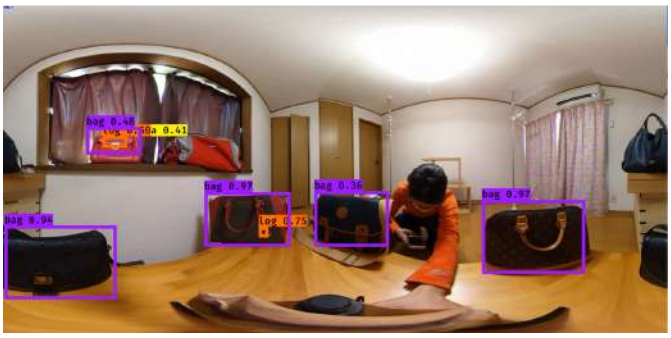

図 7　出力動画 1

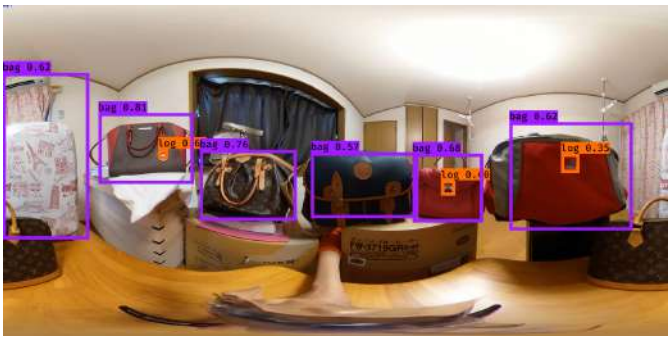

図 8　出力動画 2

5 追加実験

本章では、今回作成したモデルにより通常の画像を利用してカバン以外の物体のロゴを検出可能かどうかを確認するための実験を行った。

5.1 検出結果

入力動画は、iPhone7 により撮影された画像を利用する。今回撮影した物体は、ロゴがついた洋服と化粧品である。まとめた入力画像を図 9 に、出力画像を図 10 にを示した。この結果から、カバン以外の物体に関してもロゴ検出が可能であることが確認できた。今回教師データとして用いたカバンのロゴは、文字で表現されているものが非常に多かったため、Nivea や Pink といったロゴは精度良く検出出来ていたと思う。その一方で、入力画像の一番右のロゴのような今回の教師データにはなかったイラストによるロゴも検出できることが分かった。

図 9 入力画像

図 10 出力動画

6 まとめ

本論文では、360 度カメラで撮影された動画内におけるロゴの検出を可能にする手法を提案した。Flickr API と BUYMA を用いて取得したブランドのカバンをデータセットとしてカバンとロゴのラベル付けを行い、既存の YOLO のモデルを再学習することで物体とロゴの両方の検出を試みた。2 つの動画に対する検出結果から、検出精度の向上を目指す必要があるがロゴとカバンを同時に検出することが可能であることが示された。精度を向上させるためには、教師データを増やすことと学習の際にパラメータを変更することで対応が可能だと考えている。今回多くの画像を BUYMA から取得したが、カバンとロゴの型や種類が類似している画像が多かったために教師データとしては不十分であると考えている。また、教師データ数を増やすことでより精度が高い検出が可能であると考えている。

また、360 度カメラの特性により動画上の上部と下部には歪みが生じてしまうため、撮影時に商品を適切な場所へ配置しないと歪みの影響を受けてしまう。林田ら [8] の研究を参考にして歪んだ画像を教師データに加えて再学習することで、360 度動画中の高緯度・低緯度領域でも歪みの影響を受けずに物体検出が可能となる。　今後の課題として、上記に挙げた問題点の改善が必要である。教師データを増やすことで精度の向上を目指す。今回は Flickr や BUYMA から画像データを取得したが、カバンの形やロゴのデザインは似たものが多かったために他のサイトからも画像データを取得したいと考えている。画像データを取得する方法として、楽天株式会社が国立情報学研究所を通して公開している楽天市場のデータを利用することを検討している。また、2 点目の問題として挙げた画像内での歪みに対しては、物体の配置方法を工夫したり、歪んだ画像を教師データとして追加して再学習したりすることで対処しようと考えている。また、その他の今後の課題としては、今回はモデルの再学習のみを対象にしてロゴ検出を行ったが転移学習やファインチューニングを利用することで、精度の変化を調べて最も精度良く検出が出来るモデルを作成したい。さらに、今回はカバンのみと対象を絞っていたが、Coco Dataset も活用することで他の物体に対しても物体検出とロゴ検出を行いたい。本手法を応用することで、360 度カメラを用いた新しい形のショッピングシステムの構築と仮想店舗の実現を検討している。

文　献

[1] 楽天株式会社.”2020 年度決算短信・説明会資料”.2020-08-11. https://corp.rakuten.co.jp/investors/documents/results/

[2] 株式会社ニトリホールディングス.”IR 資料一覧”. https://www.nitorihd.co.jp/ir/library/list.html

[3] 山下利之. 企業ロゴにおけるグローバル企業のロ－カル展開. 日本感性工学会大会.2020

[4] 西本尭将, 若原 徹.Selective Search と AKAZE 特徴量を用いたロゴ検出. 情報処理学会講演論文集　第 81　回全国大会 (2019), 情報処理学会. pp.539-540. 2019

[5] 井岡良太, 三宅寿英, 前田誠一, 遠藤栄 and 馬野元秀. ニューラルネットワークによる物体認識魏技術の食品生産管理システムへの応用. 日本知能情報ファジィ学会 講演論文集, 日本知能情報ファジィ学会.pp.560-564.

[6] 藤橋一輝, 木村雅之, 金崎朝子 and 小澤順. 半教師あり学習による商品画像中の個数と位置の同時推定. 人工知能学会全国大会論文集 第 32　回全国大会 (2018), 一般社団法人 人工知能学会.,pp.11–11.

[7] 建本聡, 原田陽子 and 今井健司. 深層学習を利用したウメ「露茜」の画像による熟度分類. 農業情報研究 28(3), 108-114, 2019

[8] 林田和磨 and 横山昌平. 全天球カメラにより配信される正距円筒図法動画からのリアルタイム人物検出. 第 12 回データ工学と情報マネジメントに関するフォーラム, 2020