

物体検出のための側面を用いた訓練サンプル生成による撮影枚数の削減*

上野 高貴[†], 西山 正志^{†,††}, 岩井 儀雄^{†,††}

Reducing number of acquisition images by synthesizing training samples using aspects for object detection

Kouki UENO, Masashi NISHIYAMA, and Yoshio IWAI

We propose a novel method for synthesizing training samples to obtain high accuracy of object detection under the condition that the number of acquisition images is small. The convolutional neuronal networks for object detection require the large number of acquisition images that the angles of postures of each object are varied. Thus, it is very time-consuming to collect training samples. On the other hand, our method synthesizes training samples from the small number of aspect images that determine the variation of appearances of objects. We design how to collect aspect images based on the knowledge that there is a bias in the postures of objects. Experimental results show that our method significantly reduces the number of acquisition images while keeping high detection accuracy of a comparison method that requires the large number of acquisition images.

Key words: Object detection, Training samples, Image acquisition, Synthesis, Aspects

1. はじめに

物流システムにおける商品物体の搬入出工程は、人手による膨大な労働力を必要とする。この工程は運搬作業と仕分け作業に分けることができる。運搬作業では、物体が収納されたコンテナ型の自走式ロボット^{1,2)}が巨大倉庫において導入されつつあり、近い将来に完全自動化されると見込まれている。一方、仕分け作業では、対象となる物体の種類が非常に多いため、人手で処理を行う必要がある。高精度な物体検出の機能を持つロボットアーム^{3,4,5)}を用いることで、仕分け作業を自動化することが強く求められている。本論文では、物体を精度よく検出するために、カメラで撮影した画像をどのように用いるかについて議論する。

既に実用化されているファクトリーオートメーション分野では、物体姿勢に対して強い制約を置くことで高い検出精度を得ている。例えばベルトコンベアを流れる物体を選別する用途など、同じ物体が同じ姿勢で置かれている状況に既存手法⁶⁾は向いている。しかし、本論文で議論する仕分け作業では、複数種類の物体がコンテナ内に制約のない姿勢で配置されているため、既存手法⁶⁾を単純には適用できない。ロボットアームで把持する物体を決定するために、画像中の各物体の種類が何で、それらの物体がどの位置に存在するかを正しく検出する手法が必要となる。

一般物体認識の分野では、畳み込みニューラルネットワーク^{7,8)}を用いて物体種類を正しく見分ける手法が提案されている。さらに物体検出の分野では、物体の種類と位置とを高精度でかつ高速に検出する手法^{9,10,11)}が提案されている。また、ロボットアームで把持するポイントと制御パラメータについて自己学習を行う手法¹²⁾が提案されている。これらの既存手法を応用することで、仕分け作業を自動化することが盛んに進めら

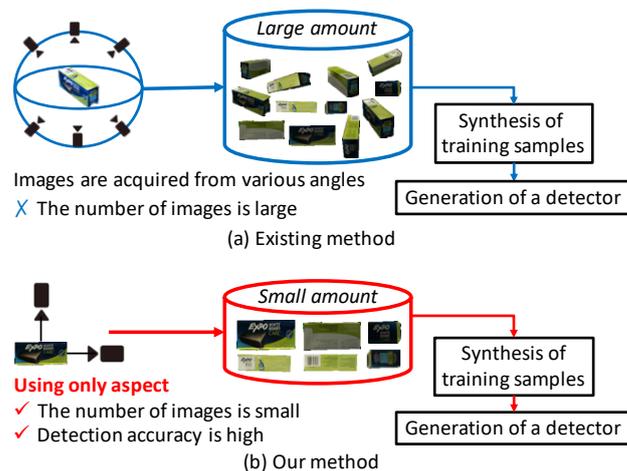


Fig.1 The purpose of our method.

れている。

畳み込みニューラルネットワークを用いた検出手法^{9,10,11)}で高い精度を得るためには、対象となる各物体について大量の訓練サンプルが必要となる。物体の姿勢が変わると見え方が大きく変動するためであり、それらの変動を十分に含む訓練サンプルを既存手法^{9,10,11)}は事前に収集しておくことを前提条件としている。図1(a)のように、物体毎にカメラ角度を密に変化させながら撮影することが一般的に多い。例えば角度を5度刻みで撮影して訓練サンプルを収集する場合、回転の3自由度を考えると $(360/5)^3 = 373248$ 枚の撮影が物体毎に要求される。訓練サンプルの収集は非常に手間がかかるため、仕分け作業の自動化にとって大きな阻害要因の一つとなっていた。

そこで本論文では、撮影枚数を大幅に減らしたとしても、高い検出精度を維持できる訓練サンプルの生成手法について述べる。コンテナ内に物体を配置した際、ありとあらゆる姿勢が発生する訳ではなく、起こりうる姿勢には偏りがあることを利用する。提案手法では、検出に有効な特定の姿勢に対する物体側面のみを収集することで、撮影枚数を減らすことを狙う。提案

* 原稿受付 平成 30 年 5 月 7 日

[†] 鳥取大学 大学院 持続性社会創生科学研究科

^{††} 鳥取大学 工学部附属クロス情報科学研究センター

手法の狙いを図 1 (b) に示す。実験結果より、撮影枚数を大幅に削減したとしても、カメラ角度を密に変化させて撮影する比較手法と同等以上の検出精度を実現できることを確認した。以下では、2. 章で物体側面について考察し、3. 章で提案手法の流れについて述べる。次に 4. 章および 5. 章で提案手法の性能を評価し、最後に 6. 章でまとめる。

2. 物体側面

2.1 概要

本論文では、一般消費者向けの商品物体を取り扱う物流システム^{1,2,13)}を想定し、その中でも仕分け作業を対象とする。この作業では、自走するコンテナから商品物体を作業員が取り出す。以下では、コンテナから商品物体を取り出す作業を、ロボットアームによる把持で自動化することを考える。把持の対象となる物体をロボットアームが正しく認識するために、コンテナ内の物体をカメラで撮影した画像から検出する。物体検出で考慮すべき点として、物体の種類と配置が挙げられる。以下でそれぞれの詳細を説明する。

2.2 物体の種類

把持作業の対象となる物体は、模様のある箱や袋に変形しないよう梱包されている場合が多い。提案手法では、物体表面は複数の面で構成されていると仮定する。ここで、ある面を床に接地させた時、その物体が転倒せず静止している状態を自立と呼ぶ。自立する面の数に応じて、物体を平面、直方体、円柱の 3 つの形状に分類する。以下、それぞれの形状について説明する。

平面 自立する面が 2 個で構成される物体とする。例：絵本、カレンダー、鉛筆、DVD、皿、ノート

直方体 自立する面が 6 個で構成される物体とする。例：石鹸、スポンジ、ティッシュ

円柱 自立する面が 2 個存在し、曲面が一部含まれる物体とする。例：テニスボール容器、ダンベル、ガムテープ

なお、上記の形状に当てはまらない物体として衣服やテトラパックなどが存在するが、本論文で対象とする仕分け作業では取り扱わないものとする。

2.3 物体の配置

把持作業の対象となる物体は、ロボットアームが接触し易い状態でコンテナ内に重なりがなく配置されることが望ましい。しかし、コンテナ容量の制約、アーム稼働範囲の制約、把持作業の繰り返しにより、物体同士が重なって配置される場合が多く発生する。複数の物体が重なって配置された直後は、それら物体の姿勢は様々に変化する。ただし、物体の姿勢は常に変化し続ける訳ではなく、図 2 (a) のように、転倒しないよう静止状態に落ち着く。このように、コンテナ内で観測される物体姿勢は、全方向へは自由に発生せず、特定の方向へ偏りが生じる。静止状態の見え方は、図 2 (b) のように、各物体において一個の面で主にカバーされていると考えられる。そこで提案手法では、物体を構成する面の見え方変動を検出器の訓練サンプルに含めることで、物体同士が重なり傾いて配置された場合であっても正しく検出することを狙う。なお本論文では、部分的に隠れている物体を検出することは目的とせず、カメラから見て最も手前にある物体を正しく検出することを考える。

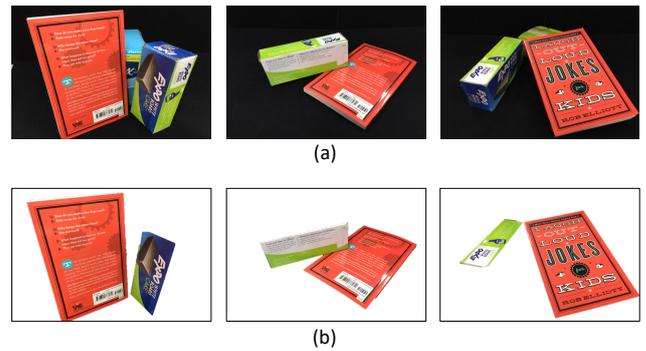


Fig.2 Main surfaces of objects observed from a camera.



Fig.3 Example of aspects of a planar object.



Fig.4 Example of aspects of a rectangular object.

2.4 側面の決め方

以下では、訓練サンプルに含める面を側面と呼び a_i ($i \in \{1, \dots, N\}$) と表記する。 N は側面数とする。2.2 節で述べた平面、直方体、円柱の物体について側面の決め方を述べる。

平面物体の側面 $a_i \in \{a_1, a_2\}$ を以下の手順で決定する。

1. 自立する 2 個の面の内で面積が大きい方の面を a_1 とする。ただし、面積が同じ面が存在する場合はランダムに 1 個を選んで a_1 とする。
2. a_1 の裏となる面を a_2 とする。

図 3 に平面物体の側面の例を示す。

直方体物体の側面 $a_i \in \{a_1, \dots, a_6\}$ を以下の手順で決定する。

1. 自立する 6 個の面の内で面積が最大の面を a_1 とする。ただし、最大面積が同じ面が複数存在する場合は、その中からランダムに 1 個を選んで a_1 とする。
2. a_1 の裏となる面を a_2 とする。
3. a_1, a_2 以外の自立する面の内で面積が最大となる面を a_3 とする。ただし、最大面積が同じ面が複数存在する場合は、その中からランダムに 1 個を選んで a_3 とする。
4. a_3 の裏となる面を a_4 とする。
5. 残り 2 個の面の内で面積が大きい方の面を a_5 とする。ただし、面積が同じ面が存在する場合は、その中からランダムに 1 個を選んで a_5 とする。
6. a_5 の裏となる面を a_6 とする。

図 4 に直方体物体の側面の例を示す。

円柱物体において底面以外は曲面であるため、図 5 のように、外接する正 P 角柱で近似する。提案手法では、近似した正 P 角柱のそれぞれの面を仮側面と呼ぶ。円柱物体は、2 個の底

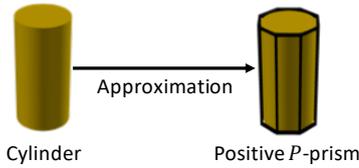


Fig.5 Approximation of a cylinder object using a positive P -prism.

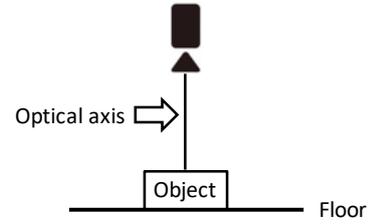


Fig.8 The relationship of positions between a camera and a target object.

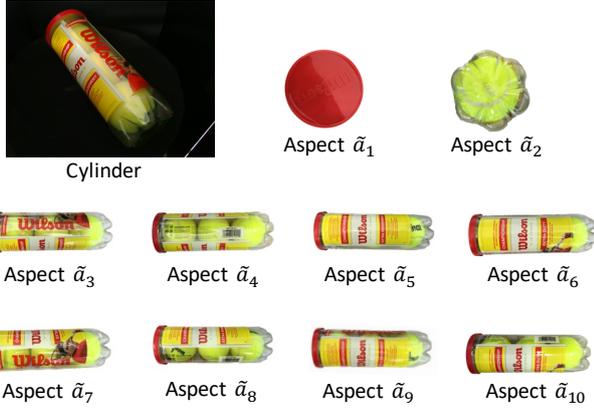


Fig.6 Example of aspects of a cylinder object.

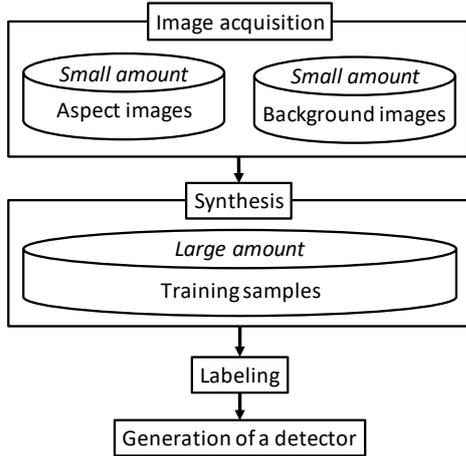


Fig.7 Overview of our method.

面を合わせて $P + 2$ 個の仮側面で構成されると仮定する。円柱物体の仮側面 $\tilde{a}_i \in \{\tilde{a}_1, \dots, \tilde{a}_{P+2}\}$ を以下の手順で決定する。

1. 底面 2 個の中からランダムに 1 個を選び \tilde{a}_1 とする。
2. \tilde{a}_1 の裏となる面を \tilde{a}_2 とする。
3. 底面以外の面からランダムに 1 個を選んで、 \tilde{a}_3 とする。
4. \tilde{a}_3 に隣接する面を時計回りに \tilde{a}_4 から \tilde{a}_{P+2} とする。

図 6 に正 8 角柱で近似した場合の円柱物体の仮側面の例を示す。

3. 提案手法

3.1 概要

物体側面の考え方を提案手法に導入することで、訓練サンプル収集時の撮影枚数を削減することを考える。提案手法の流れを図 7 に示す。まず、側面画像と背景画像を撮影する。次に、側面画像と背景画像から訓練サンプルを仮想的に生成する。生成した訓練サンプルにラベルを付与し、検出器を学習させる。以下、提案手法の詳細を述べる。

3.2 画像撮影

側面画像の撮影手法について述べる。カメラの光軸が床面に対して垂直になるようにカメラを設置し、対象物体を自立させた状態で床面に配置する。その際、カメラの光軸が物体重心を通るものとする。図 8 にカメラと物体の位置関係を示す。各物体の撮影枚数は 2.4 節で定めた側面の数とする。平面物体は $\{a_1, a_2\}$ の 2 枚、直方体物体は $\{a_1, \dots, a_6\}$ の 6 枚、円柱物体は $\{\tilde{a}_1, \dots, \tilde{a}_{P+2}\}$ の $P + 2$ 枚を準備する。なお、物体の側面領域はセグメンテーションされているものとし、側面領域の外接矩形を側面画像のサイズ $W_a \times H_a$ 画素とする。

次に、背景画像の撮影手法について述べる。訓練サンプルの背景を検出時と同じ見え方とするために、物体を格納するコンテナを撮影する。コンテナ内に物体が存在しない場合、および、物体が存在する場合を撮影の対象とする。ただし、背景画像内に存在する物体は、検出対象の物体には含まれないものとする。背景画像のサイズを $W_b \times H_b$ 画素とする。

3.3 側面を用いた訓練サンプル生成

訓練サンプルの見かけの数を増加させる手法として、一般的にはデータ拡張¹⁴⁾が用いられる。データ拡張では、撮影された 1 枚の物体画像に対して、回転、拡大縮小、平行移動、彩度、明度について画像変換を施す。複数の画像を仮想的に生成することで、訓練サンプル数を増加させる。提案手法では、物体が側面で構成されることから、射影変換を従来のデータ拡張の手法に加え適用する。側面が斜めに配置された見え方を射影変換で仮想的に生成することで、物体同士の重なりにより傾いて配置された場合であっても検出精度を向上させることを狙う。さらに、複数の側面画像を背景画像に重畳することにより、コンテナ内での物体の配置位置の違い、物体同士の重なる順序の違いに対応することを狙う。

訓練サンプル生成について、その流れを図 9 に示し、手順を以下で述べる。

1. 背景画像を 1 枚ランダムに選択する。
2. 側面画像を 1 枚ランダムに選択する。
3. 側面画像に対して射影変換、回転、拡大縮小、平行移動を順に適用する。
4. 背景画像に側面画像を重畳する。
5. 一定の確率 p で手順 2 に戻る。なお p は物体を追加するかどうかの確率とする。
6. 画像全体に対して彩度、明度を変更する。

この手順を繰り返すことで、 T 枚の訓練サンプルを仮想的に生成する。射影変換、回転、拡大縮小、平行移動、彩度、明度の変換に用いるパラメータを以下に挙げる。

射影変換 側面画像の端の四点に対して、変換後の端の四点を $(W_a/E, 0), ((1 - 1/E)W_a, 0), (0, H_a), (W_a, H_a)$ とし射影する。

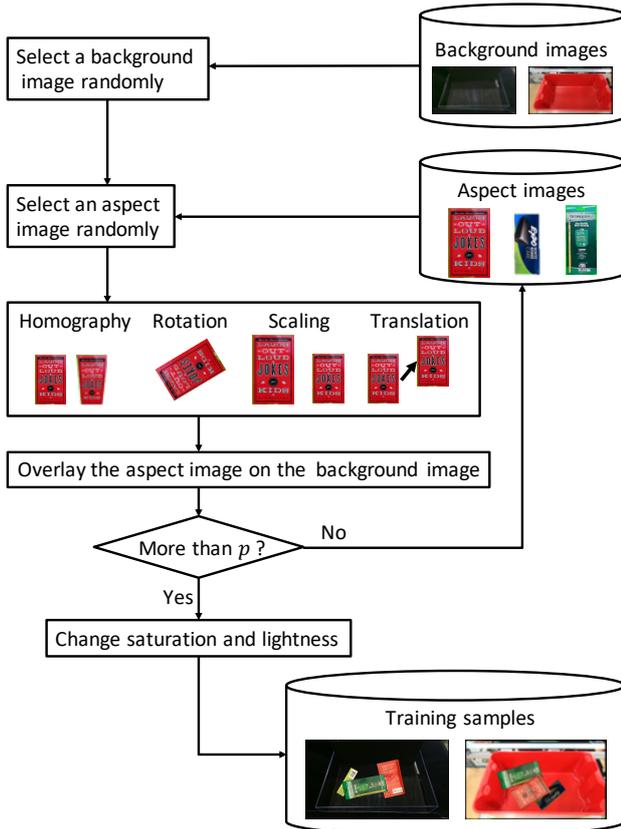


Fig.9 Overview of synthesizing training samples.

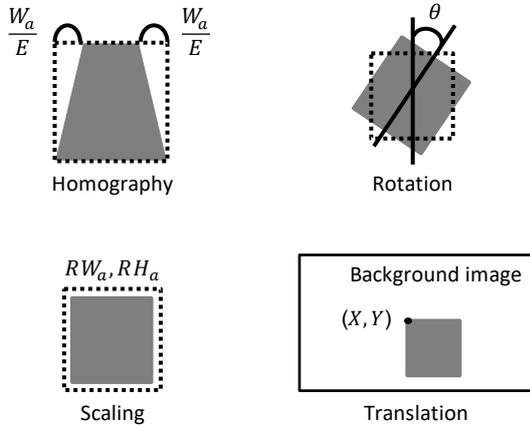


Fig.10 Parameters for synthesizing training samples.

回転 側面画像の中心を原点とし θ 度だけ回転する。

拡大縮小 側面画像を R 倍する。

平行移動 背景画像中の位置 (X, Y) に側面領域を重畳する。

彩度 重畳後の訓練サンプルの彩度を HSV 色空間で S だけ加算する。

明度 同様に明度を V だけ加算する。

これらのパラメータの一部を図 10 に示す。

3.4 訓練サンプルのラベル付け

一般的に、畳み込みニューラルネットワークを用いた検出手法^{9, 10, 11)}では、物体の種類数そのままクラス数となる。この場合、ある物体の様々な姿勢に対する見え方変動を、単一クラスの訓練サンプルに含めることになる。よって、同一クラス内における見え方変動が非常に大きくなり、ニューラルネット

Yellow : Planar, Red : Rectangular, Blue : Cylinder



Fig.11 Target objects for evaluating basic performance.

ワークの構造を深く設計したとしても、検出精度が低下する場合があります。そこで提案手法では、種類クラスに加えて、側面クラスをラベルとして各訓練サンプルに与える。これにより、同一クラスに属する見え方変動が小さくなり、検出精度を向上させることができると考えられる。具体的には2.4節で定めた側面毎に1つのクラスを与える。平面物体は1種類につき $\{a_1, a_2\}$ の2クラス、直方体物体は1種類につき $\{a_1, \dots, a_6\}$ の6クラス、円柱物体は1種類につき $\{\tilde{a}_1, \dots, \tilde{a}_{P+2}\}$ の $P+2$ クラスとする。

4. 基本性能の評価

4.1 実験条件

物体側面を用いた訓練サンプル生成の有効性を確認するために24種類の物体を用いた検出精度の評価を行った。図11に対象物体を示す。平面形状が13種類、直方体形状が7種類、円柱形状が4種類とした。なお円柱物体は $P=8$ で近似した。

テストセットとして、物体同士の重なりがないコンテナ、および、物体同士の重なりがあるコンテナの2タイプを設定した。図12(a)のテストセット1では物体の重なりがない条件、(b)のテストセット2では物体の重なりがある条件で、それぞれ画像を撮影した。サンプル数は、テストセット1が272枚、テストセット2が68枚とした。

提案手法の側面画像を収集するために、24種類の対象物体に対して合計で $13 \times 2 + 7 \times 6 + 4 \times 10 = 108$ 枚の撮影を行った。また背景画像を収集するために、テストセット1用に5枚、テストセット2用に4枚の撮影を行った。側面画像の平均解像度を $W_a = 133.0 \pm 68.5$ 画素、 $H_a = 110.2 \pm 59.7$ 画素とし、背景画像の解像度を $W_b = 640$ 画素、 $H_b = 360$ 画素とした。

訓練サンプルを3.3節の手法で仮想的に生成した。生成時のパラメータは一樣乱数で設定し、その範囲は E を $[4, 6]$, R を $[0.99, 1.01]$, θ を $[-180, 180]$, X を $[0.2W_b, 0.6W_b]$, Y を $[0.2H_b, 0.6H_b]$, S を $[-0.07, 0.07]$, V を $[-0.11, 0.11]$ とした。生成した訓練サンプルとして、図13(a)にテストセット1に対する例を、(b)にテストセット2に対する例を示す。

検出器としてYOLOv2¹¹⁾を使用した。バッチサイズは64とした。Learning rateは0.001からスタートして学習が進むにつれ0.0001, 0.00001と1/10ずつ小さくした。ネットワーク構造として、畳み込み層を23個、プーリング層を5個、全結合層を1個を用いた。



Fig.12 Test sets for evaluating detection accuracy.



Fig.13 Synthesized training samples for generating a detector.

検出精度の指標として intersection over union (IOU) を用いた。IOU は、各物体について検出された矩形領域と教示された矩形領域の論理積および論理和で定義される。図 14 に IOU の例を示す。IOU が 0.5 以上である場合を正しく検出できたと判定した。テストセット毎に各サンプル中の各物体の検出率を算出し、その平均と標準偏差を算出した。

各テストセットにおいて物体同士に重なりが発生した場合、カメラから見て最も手前に存在する物体を検出率の算出対象とした。この場合、1 個の物体に対して複数の面が同時にカメラから観測される例が多く発生していた。検出された矩形領域、および、教示された矩形領域の例を図 15 に示す。なお、物体の外接矩形を手作業で教示した。1 個の物体に対して複数の面が同時にカメラから観測された画像であっても、検出

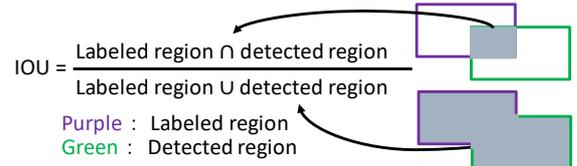


Fig.14 Example of IOU.



Fig.15 Example of detected region and labeled region. The purple box represents labeled region. The green box represents detected region.

された矩形領域は 1 個であることが多かった。これは、畳み込みニューラルネットワークから出力された複数の重なり合う矩形領域が、YOLOv2 内に実装されている Non-Maximum Suppression (NMS) 処理によって 1 個に統合されていたためである。検出された矩形領域と教示された矩形領域との間で IOU を計算し 0.5 以上であれば正解とした。NMS 処理後において 1 個の物体に対して複数の矩形領域が検出された場合は、IOU の計算を繰り返した。

4.2 様々な角度から撮影した訓練サンプルとの比較

提案手法により撮影枚数を減らしたとしても同等の検出精度を得ることができるかどうか検証した。提案手法では 4.1 節で述べた 108 枚の側面画像を用いた。ここで比較に用いる手法では、各物体について姿勢を 15 度刻みに変化させて撮影した 16176 枚の物体画像を用いた。以下では比較手法 A と呼ぶ。各物体の撮影枚数以外の実験条件は、提案手法、比較手法 A とともに全て 4.1 節と同じとした。なお訓練サンプルは両者とも 3.3 節の手法で生成し、その数は両者とも $T = 20$ 万枚とした。

図 16 (a) でテストセット 1 について、(b) でテストセット 2 について検出精度を示す。図中では、全物体、および、物体形状毎の検出率の平均と標準偏差を記載した。どちらのデータセットでも、提案手法は比較手法 A と比べて良好な検出精度であった。実験結果より、提案手法を用いることで撮影枚数を 1/149 に削減したとしても削減前と同等以上の精度を得ることができることを確認した。

4.3 側面をランダムに選択した場合との比較

提案手法により側面を選択した場合と、ランダムに側面を選択した場合で検出精度を比較した。ここで比較に用いる手法では、4.2 節で用いた 16176 枚の物体画像の中からランダムに側面画像を設定した。以下では比較手法 B と呼ぶ。側面の選択以外の実験条件は、提案手法、比較手法 B とともに全て 4.1 節と同じとした。側面画像の合計枚数は両者とも 108 枚とした。

図 17 (a) でテストセット 1 について、(b) でテストセット 2 に

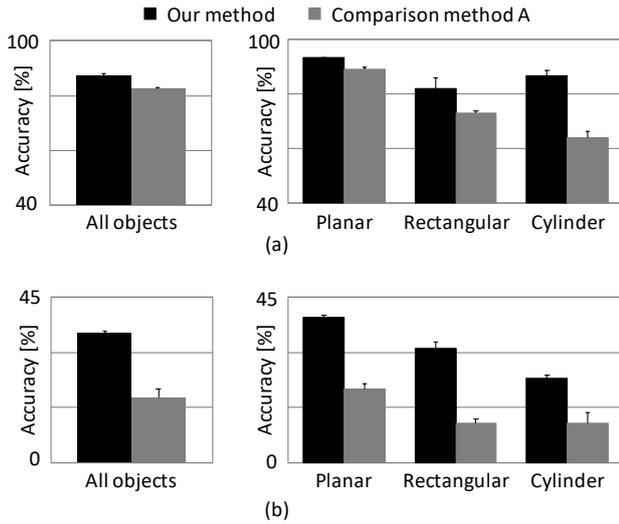


Fig.16 Detection accuracy using our method or the comparison method A.

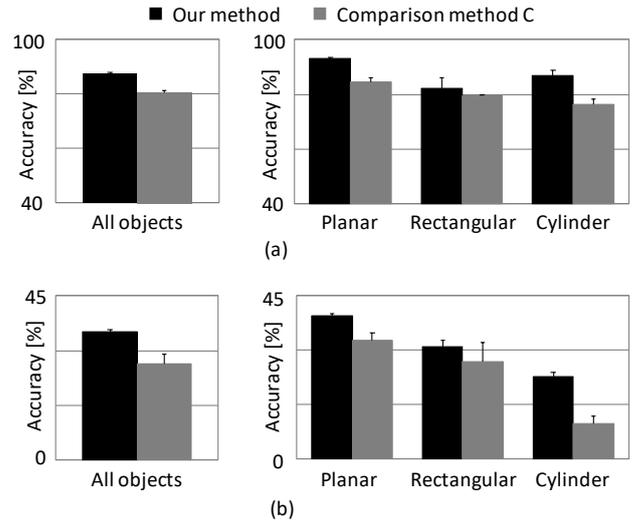


Fig.18 Detection accuracy using our method or the comparison method C.

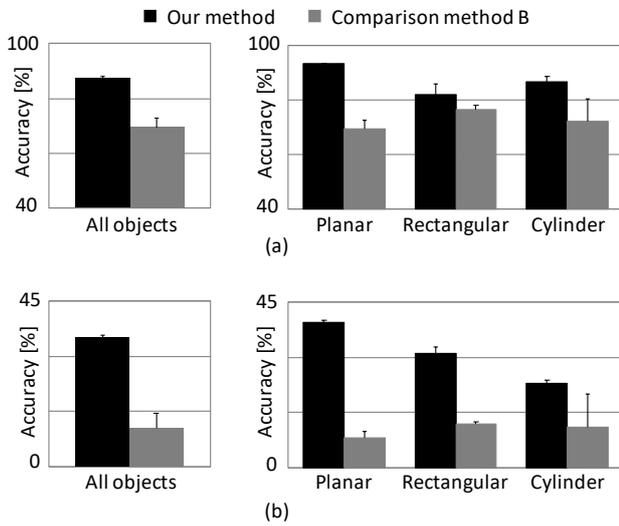


Fig.17 Detection accuracy using our method or the comparison method B.

ついて検出精度を示す。提案手法は比較手法 B と比べて高い検出精度を得ていた。実験結果より、2.4 節で述べた側面の選択手法はランダム選択と比べて有効であることが確認できた。

4.4 側面クラスを加えた場合の有効性の検証

物体の種類クラスに加えて側面クラスを用いる場合の検出精度を比較した。提案手法では、3.4 節で述べたように種類クラスに側面クラスを加えて検出器を学習した。ここで比較に用いる手法では、種類クラスのみを用いて検出器を学習した。以下では比較手法 C と呼ぶ。提案手法の総クラス数は 108、比較手法 C の総クラス数は 24 とした。クラスの分け方以外の実験条件は、提案手法、比較手法 C とともに全て 4.1 節と同じとした。検出精度は両者ともに、ここまでの実験と同様、推定された種類と領域位置から IOU で算出した。

図 18 (a) でテストセット 1 について、(b) でテストセット 2 について検出精度を示す。提案手法は比較手法 C と比べて高い検出精度を得た。訓練サンプル数は同じであっても、検出器の学習時に種類クラスと側面毎クラスとを同時に用いることで、同一クラス内での見え方の変動が抑えられ検出精度が向上したと考えられる。

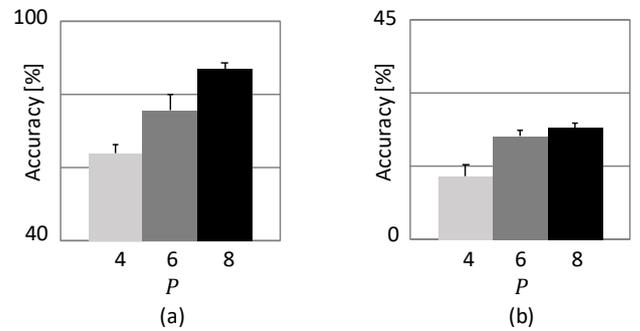


Fig.19 Detection accuracy while changing the parameter of the positive P -prism.

4.5 正 P 角柱のパラメータを変更した場合の評価

円柱を多角柱で近似する際のパラメータ P を変更した場合の検出精度を評価した。ここでは $P = 4, 6, 8$ について比較した。側面画像の数はそれぞれ 92, 100, 108 枚とした。図 19 (a) でテストセット 1 について、(b) でテストセット 2 について検出精度の変化を示す。実験結果より、円柱を多角柱で近似する際、仮側面の数を増加させた方が検出精度が高まることが分かった。

4.6 訓練サンプル生成時の射影変換の有効性の検証

側面を用いた訓練サンプル生成において、射影変換を適用する場合、および、適用しない場合の精度を比較した。射影変換を適用しない場合を以下では比較手法 D と呼ぶ。図 20 (a) でテストセット 1 について、(b) でテストセット 2 について検出精度を示す。

提案手法は比較手法 D と比べてテストセット 1 において検出精度が改善されていた。この理由を考察する。テストセット 1 の画像を撮影した際、カメラの光軸と床の法線は一致しておらず、射影による見え方変化が発生していた。この見え方変化を射影変換によって訓練サンプルに含めることができたため、精度が向上したと考えられる。また、テストセット 2 においても、提案手法は比較手法 D と比べて高い検出精度を得ていた。テストセット 2 では物体同士の重なりが存在するため、射影による見え方変化がテストセット 1 と比べてより大きく発生して

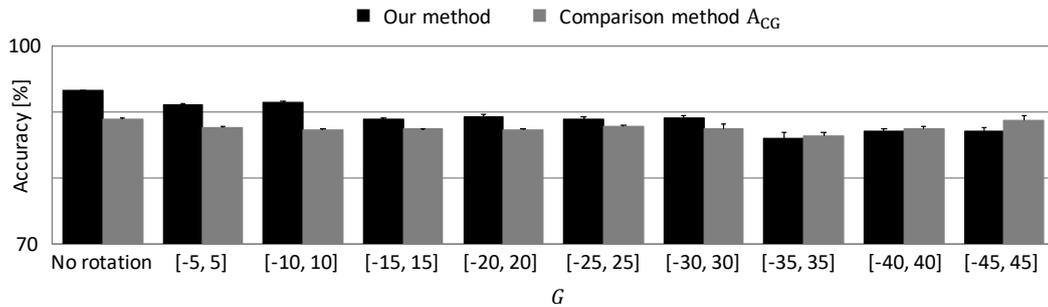


Fig.25 Detection accuracy using our method or the comparison method A_{CG} .



Fig.24 Examples of CG training samples.

回転摂動について $G = 0$ から $G = 45$ まで 5 度刻みで CG テストセットを作成した。図 25 で CG テストセットにおける検出精度を示す。提案手法は比較手法 A_{CG} と比べて、30 度以下では良好な検出精度であることが分かった。また、 G が大きくなるにつれ、提案手法と比較手法 A_{CG} との間で精度差は小さくなり、35 度を超えると精度が逆転することが分かった。回転摂動の範囲を 20 度までと想定し訓練サンプルを作成した場合、テストセットの回転摂動の範囲は 30 度まで対応できることが分かった。また提案手法を用いることで、撮影枚数を 1/108 に削減したとしても、様々な角度から撮影した場合と同等以上の精度を得られることを確認した。

6. おわりに

本論文では、少ない撮影枚数でも高い検出精度を得ることができる訓練サンプルの生成手法について述べた。提案手法では、コンテナ内に複数物体が配置される際、それらの物体が取り得る姿勢に偏りが存在することに着目し、物体側面の考え方を導入した。物体側面の見え方変動を仮想的に生成することで、検出器の訓練サンプルに多様性を持たせた。実験結果により、撮影枚数を大幅に削減したとしても、様々な姿勢を撮影する比較手法と同等以上の検出精度が維持できることを確認した。

本論文で対象とした物流システムの仕分け作業以外にも、物体が取り得る姿勢に偏りが発生する事例は複数存在する。例えば、店舗におけるレジ打ち作業、店舗における棚割り作業、家庭における食品管理作業が想定される。今後の課題として、様々な事例を想定したデータセットでの検出精度の評価、物体

種類や物体形状の数が増加した時の検出精度の評価、物体の種類と位置に加えて姿勢も同時に推定する手法の開発などが挙げられる。

謝 辞

本研究の一部は、人工知能研究振興財団の助成を受けたものである。評価データセットの一部は東芝インフラシステムズ株式会社の下山賢一氏から提供して頂いた。

参 考 文 献

- 1) P. R. Wurman, R. D' Andrea, and M. Mountz. Coordinating hundreds of cooperative, autonomous vehicles in warehouses. In *Proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence*, pp. 1752–1759, 2007.
- 2) R. D' Andrea. Guest editorial: A revolution in the warehouse: A retrospective on kiva systems and the grand challenges ahead. *IEEE Transactions on Automation Science and Engineering*, Vol. 9, No. 4, pp. 638–639, 2012.
- 3) P. Baker and Z. Halim. An exploration of warehouse automation implementations: cost, service and flexibility issues. *Supply Chain Management: An International Journal*, Vol. 12, No. 2, pp. 129–138, 2007.
- 4) 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. *IEEE Transactions on Automation Science and Engineering*, Vol. 15, No. 1, pp. 172–188, 2018.
- 5) H. Fujiyoshi, T. Yamashita, Y. Yamauchi, T. Hasegawa, M. Hashimoto, S. Akizuki, Y. Domaie, and R. Kawanishi. Team c^2m : Two cooperative robots for picking and stowing in amazon picking challenge 2016. In *Proceedings of Warehouse Picking Automation Workshop*, pp. 1–4, 2017.
- 6) C. A. Rosen and D. Nilzan. Use of sensors in programmable automation. *Computer*, Vol. 10, No. 12, pp. 12–23, 1977.
- 7) A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097–1105, 2012.
- 8) K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of International Conference on Learning Representations*, pp. 1–14, 2015.
- 9) W. Liu, D. Anguelov, and D. Erhan. Ssd: Single shot multibox detector. In *Proceedings of European Conference on Computer Vision*, pp. 21–37, 2016.
- 10) S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, 2017.
- 11) J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517–6525, 2017.
- 12) P. Pastor S. Levine, A. Krizhevsky, J. Ibarz, and D. Quillen. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *International Journal of Robotics Research*, pp. 421–436, 2017.
- 13) H. Liu J. Li. Design optimization of amazon robotics. *Automation, Control and Intelligent Systems*, Vol. 4, No. 2, pp. 48–52, 2016.
- 14) A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, and T. Brox. Discriminative unsupervised feature learning with convolutional neural networks. In *Proceedings of Neural Information Processing Systems*

Conference, pp. 766–774, 2014.

- 15) C. Rennie, R. Shome, K. E. Bekris, and A. F. D. Souza. A dataset for improved rgbd-based object detection and pose estimation for warehouse pick-and-place. *IEEE Robotics and Automation Letters*, Vol. 1,

pp. 1179 – 1185, 2016.