

## CNN とステレオカメラより得た距離情報を用いた複数物体認識手法 Multiple Object Recognition Method Using CNN and Distance Information Obtained from Stereo Camera

速水健太† 土屋誠司‡ 渡部広一‡

Kenta Hayamizu Seiji Tsuchiya Hirokazu Watabe

### 1 はじめに

近年、人間の役に立つパートナーロボットが注目されている。このロボットに求められる技術の一部として、ロボットの自己位置認識、目的地までの自律移動、指示された物の認識などが挙げられる。これらの技術には、ロボットの視覚から得られる情報を必要としているため、外界の情報を認識する技術が必要不可欠である。外界の情報を認識する技術の 1 つとして、一般物体認識がある。一般物体認識とは、制約のない実世界で撮影された 1 枚の画像に写る 1 つの物体を認識して、“ライオン”や“リンゴ”のように、その物体の一般名称を出力する研究分野である。一般物体認識では、機械学習の発展とともに認識精度が向上しており、画像処理における機械学習では、畳み込みニューラルネットワーク<sup>[1]</sup> (以降 CNN とする) が一般的に使用されている。

一般物体認識における制約のない実世界の画像には、複数の物体が写っている場合も含む。複数の物体が写っている場合の一般物体認識を複数物体認識と呼ぶことにする。図 1 に複数物体認識の例を示す。先行研究<sup>[2]</sup>では、CNN と 3 つの物体抽出法と輪郭取得手法を用いて複数物体認識を行っている。その複数物体認識は単にテスト用画像として撮影した画像を対象に行っているにすぎない。実際にロボットの視覚 (ステレオカメラ) から取得できる画像で判別しなければ実用性がないと考えられる。また、本研究室で Robovie-RVer3 (以降 Robovie とする) を用いての先行研究では、ステレオカメラで周囲の環境を把握する際に色抽出でのみ判別を行っている。

そこで、本研究では、先行研究の技術をロボットに搭載されているステレオカメラから抽出できる画像を用いて複数物体認識を行う。それに加え、ステレオマッチングにより得ることができる距離情報を用いた複数物体認識システムを提案する。



図 1 複数物体認識の例

### 2 関連研究

#### 2.1 CNN

CNN は畳み込みとプーリングの 2 つの計算を交互に繰り返す順伝播型のネットワークである。本稿では、Alexnet<sup>[3]</sup>と呼ばれるネットワーク構造を使用する。Alexnet のネットワーク構造を図 2 に示す。

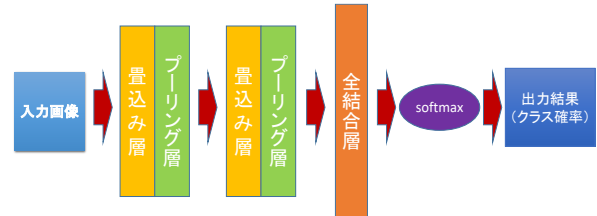


図 2 Alexnet のネットワーク構造

入力画像に対するフィルタの畳み込み処理を行う層と、その出力に対するプーリング処理を行う層を交互に複数回繰り返す。そして、全結合層の出力に対して softmax を適用し、入力画像に対するクラス確率を最終的な出力とする。なお、図 2 中の conv, pool, norm, fc はそれぞれ畳み込み層、プーリング層、正規化層、全結合層を表す。

CNN を構築する際の学習パラメータについて述べる。本稿で使用する学習データは、30 種類の物体、総画像数 37,784 枚を使用し、学習は 100,000 回行った。

### 3 提案手法

#### 3.1 概要

提案する複数物体認識システムの流れを図 3 に示す。

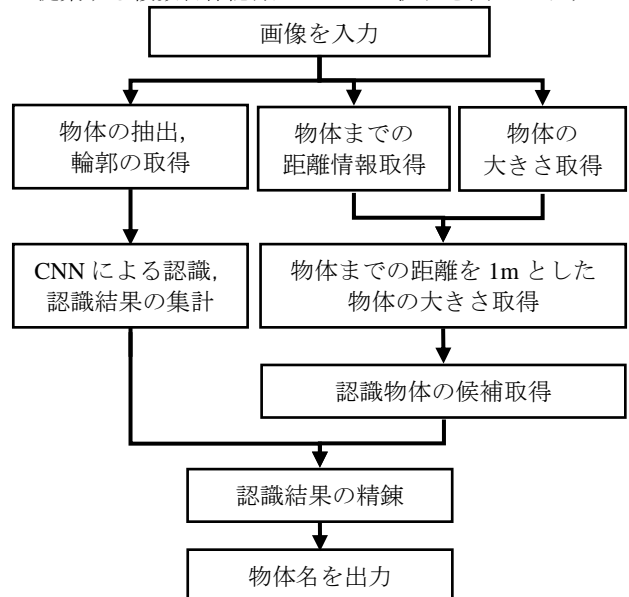


図 3 提案システムの流れ

#### 3.2 物体の抽出, 輪郭の取得

CNN による認識の前処理として、SURF 特徴点、色情報、画像ピラミッドを用いた 3 通りの物体抽出手法で物体抽出を行っていた。その後、物体抽出された各画像に対して、物体の位置に依存しない輪郭取得手法を用いて得た輪郭画像を CNN で認識する。

#### 3.3 CNN による認識, 認識結果の集計

各画像から得られた認識結果から、物体名と確率を集計し、最終的な物体名を出力する。

† 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

‡ 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

### 3.4 物体までの距離情報取得

Rovobie の左右のステレオカメラから取得した 2 枚の画像にキャリブレーション<sup>4)</sup>を施す。Rovobie の左右のステレオカメラには自由度があり、平行ステレオではないので、ステレオマッチングを行う際、2 枚の画像の対応点の探索が難しく、適切な結果が得られない。そのため、キャリブレーションを施し、画像を補正する。キャリブレーション後、ステレオマッチング<sup>5)</sup>を行い、視差画像を取得する。取得した視差画像から物体までの距離を取得する。

### 3.5 物体の大きさ取得

Rovobie の左ステレオカメラから取得した画像にエッジ抽出<sup>6)</sup>を施す。エッジ抽出後の画像から物体部分を抽出し、その抽出した画像の面積を物体の大きさとして取得する。

### 3.6 物体までの距離を統一した物体の大きさ取得

物体の大きさを比較する際、物体毎に物体までの距離が異なると比較できない。物体は、近くにあればあるほど大きく見え、遠くにあればあるほど小さく見えるからである。そこで、物体の大きさを比較するには、物体までの距離を統一する必要がある。その際、物体の大きさは、距離の 2 乗に反比例する関係<sup>7)</sup>を用いる。取得した物体の大きさと物体までの距離の 2 乗の積で、物体までの距離を 1m とした物体の大きさを取得する。

### 3.7 認識物体の候補取得

CNN で使用する物体クラスそれぞれに大きさを設定する。それと物体までの距離を 1m とした物体の大きさを比較し、差の小さい上位 3 つの物体を候補として取得する。

## 4 提案システムの評価

提案したシステムを評価するために、表 1 の評価基準に沿って評価する。提案システムにおける評価結果を図 4、図 5 に示す。評価に使用するテスト画像は、Rovobie の左カメラから取得した 2 物体画像 30 枚、3 物体画像 30 枚の合計 60 枚である。Rovobie の右カメラから取得した画像は、物体までの距離情報取得にのみ使用している。

表 1 提案システムの評価基準

◎ : 正解の物体名が上位から順に全て出力された
○ : 正解の物体名が全て出力された
△ : 少なくとも 1 つ正解の物体名が出力された
× : 正解の物体名が 1 つも出力されなかった

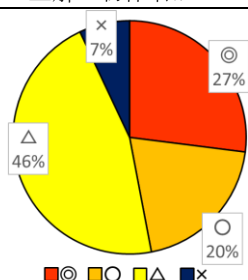


図 4 2 物体の評価結果

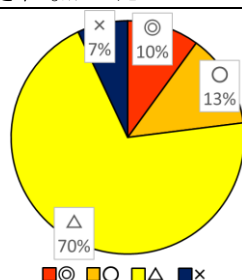


図 5 3 物体の評価結果

## 5 考察

CNN による認識について、本研究では、ステレオカメラで撮影した画像を用いて評価を行った。しかし、CNN の学習画像はステレオカメラで撮影したものを使用していない。そのため、物体をうまく抽出できたとしても、CNN で認識できない、認識しても認識率が低い結果が得られたと考えられる。ステレオカメラで撮影した画像を CNN の学習画像に加えることで、解決できるのではないかと考えられる。

物体までの距離を取得する際、物体毎に距離を取得していない。Robovie から 2 物体あるいは 3 物体の平均距離を取得している。2 物体、3 物体がロボットからある程度同じ距離にあれば問題はない。しかし、物体毎にロボットからの距離が異なると認識物体の候補をうまく取得できない。また取得した距離に 10~20cm の誤差が生じ、認識物体の候補を取得する際、正しい候補が取得できない場合もみられた。より良い視差画像から実距離の取得する手法、キャリブレーション法、ステレオマッチング法を用いる事で誤差が小さくなるのではないかと考えられる。

物体の大きさ取得について、図 6 に原画像、図 7 に物体の大きさ取得のための物体抽出画像、図 8 に理想の物体抽出画像を示す。



図 6 原画像

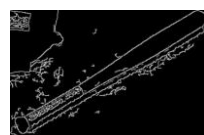


図 7 取得結果



図 8 取得結果(理想)

図 7 の結果では、バットは抽出できているが、取得した画像の面積を物体の大きさとして取得しているため、バットより大きい大きさを取得してしまう。図 8 のようなバット部分のみを抽出できるよう、物体の配置の仕方考慮した物体抽出を行うことが必要であると考えられる。また、物体が画像に対して小さすぎると、物体抽出画像が出力されない結果も見られた。一部を拡大した画像も使用することや、より柔軟な輪郭の大きさの設定を行うことが必要であると考えられる。

## 6 おわりに

本研究では、ロボットのステレオカメラから取得した画像を使用し、CNN とステレオマッチングにより得られる距離情報を利用した複数物体認識システムを構築した。しかし、問題点が複数あり、理想的な物体認識が行えたとは言うことはできない。より良いステレオマッチング手法、視差画像から実距離への変換手法、物体の大きさ取得手法の構築により精度の高い実用的な複数物体認識の実現に近づけるのではないかと考える。

### 謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けた。

### 参考文献

- [1] 山下隆義, "イラストで学ぶディープラーニング", 講談社, 2016.
- [2] 柳耀成, "畳み込みニューラルネットワークを用いた複数物体認識における物体抽出手法の提案", 同志社大学卒業論文, 2018.
- [3] Krizhevsky, Alex, Ilya, Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Advances in neural information processing systems, pp.1097-1105, 2012.
- [4] OpenCV2 プログラミングブック製作チーム, OpenCV2 プログラミングブック, 株式会社マイナビ, pp.177-191, 2011.
- [5] "opencv.jp - OpenCV-1.0: サポートされない実験的な関数", [http://opencv.jp/opencv-1.0.0/document/opencvref\\_cvaucl.html#dcl\\_cvFindStereoCorrespondence](http://opencv.jp/opencv-1.0.0/document/opencvref_cvaucl.html#dcl_cvFindStereoCorrespondence), 9.Feb.2015.
- [6] "Canny 法によるエッジ検出 — OpenCV-Python Tutorials 1 documentation", [http://labs.eecs.tottori-u.ac.jp/sd/Member/oyamada/OpenCV/html/py\\_tutorials/py\\_imgproc/py\\_canny/py\\_canny.html](http://labs.eecs.tottori-u.ac.jp/sd/Member/oyamada/OpenCV/html/py_tutorials/py_imgproc/py_canny/py_canny.html).
- [7] "遠近感 - 概要 - パースフリークス", <http://www.persfreaks.jp/main/intro/pers/>.