

ユーザの音声指示に対応した片付けロボットに関する研究 Research on Cleaning-up Robot Considering Voice Instructions of Users

張斌[†] 王君彦[†] 林憲玉[†]
Bin Zhang Junyan Wang Hun-ok Lim

1. はじめに

近年、日本においては、更なる少子高齢化や人口減少が予想されている。約 40 年後には 65 歳以上の人口割合が約 10% 増加し、労働力不足などの様々な社会的問題が起これらと考えられる。そこで、家庭、オフィス、コンビニエンスストアなどで、人の代わりに片付けてくれるロボットに対する期待が高まっている。片付けロボットには生活環境にある様々な物体を認識して物体を把持し、所定の場所に正確に片づけることが求められている。HSR ロボットなどが開発され、任意に置かれた物などの片付け作業に応用できると言われている。物体を認識するためには深層学習を用いた物体検出手法の YOLO がよく利用されている。カテゴリに対し、視覚だけでなく、聴覚と触覚を融合したマルチモーダル情報で高精度な分類手法が提案されている[1]。深層学習を用いた片付けに対する事前情報がある場合の物体検出の例として、Convolutional Neural Network (CNN) [2] や Fuzzy Color Histogram (FCH) [3]、You Only Look Once (YOLO) [4] がある。これらの手法は学習済みの物体であれば高精度に物体認識が可能である。しかし、これまで多くのロボットは、ユーザのリアルタイムな指示に対する片付けが難しい。

そこで、本研究では、音声認識システムを組み合わせたユーザの意向に沿った片付けシステムを開発する。ユーザが音声で片付ける物体を片付けロボットに指示すると、ロボットはその物体を検出・認識する。複数の物体の検出・認識には深層学習を用いたインスタンスセグメンテーション手法を利用している。この手法により物体の輪郭と形状を抽出し、ロボットアームの適切な把持姿勢を生成する。片付けロボットによる片付け実験を行い、片付けシステムの有効性を検証する。



Fig. 1 片付けロボット

2. 片付けロボットのハードウェア

本研究で使用した片付けロボットは、株式会社アールティの Sciurus17 という上半身人型ロボットである。Sciurus17 の外見を Fig.1 に示す。サイズは 270×393×665(mm)で、設置用固定金属を含めて重量は約 6kg である。

[†] 神奈川大学 Kanagawa University

腕は 7 自由度を持っており、作業が可能な範囲はロボットを中心として半径 0.6m である。頭部には RealSense D415 を搭載しており、ロボット周囲の RGB-D 環境情報を取得することができる。

3. 音声指示に対応した物体把持手法

片付けロボットの制御システムは Fig.2 に示す。まず、RGB-D センサ RealSense D415 により、ロボット前の環境情報を取得し、物体のカテゴリと領域検出を行う。次に、ユーザの音声指示を認識し、指示物体が検出されている場合は片付け対象として選出する。それから、物体の輪郭情報を取得し、最適な把持姿勢を生成する。Sciurus17 の把持機能は、ROS (Robot Operating System) を利用し、対象物体から近い方の腕で順次に物体を把持し、事前に指定した収納場所まで運搬する。

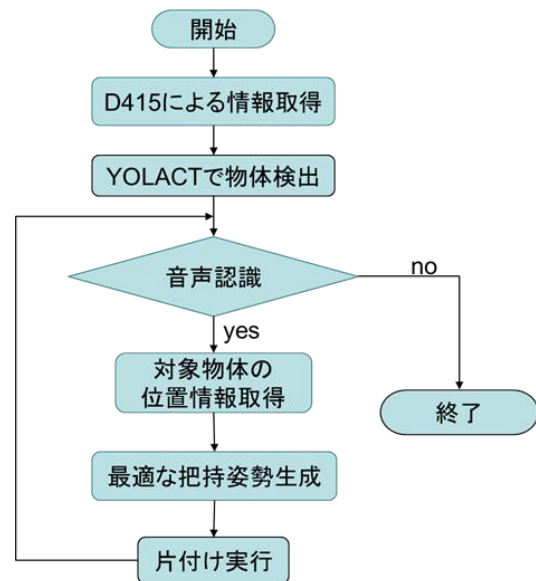


Fig. 2 処理の流れ

3.1 音声認識

Google が開発した speech-to-text API である Google speech recognition 機能を用いて、ユーザの音声を認識し、テキストに変換する。その中の名詞を検出し、事前に登録した物体対象単語とマッチングすることで、把持対象を確認する。

3.2 物体認識と物体分割

深層学習手法をベースにした YOLACT[5]を用いて、リアルタイムで物体認識を行い、物体ごとの領域を分割する。YOLACT は物体認識手法である YOLO と物体ごとに分割可能なインスタンスセグメンテーションを統合した手法である。日常生活によくみられる物体を正確に認識・分割されている。本研究で事前に登録した物体対象単語は

「bottle」、「cup」、「sports ball」と限定している。認識できない物体は未知のカテゴリにする。

3.3 把持姿勢生成

分割した物体の画像上の領域から物体の 3D 点群データを取得し、物体の中心座標を算出する。また、検出した物体のバウンディングボックス内に物体分割した部分と空白の部分で、物体が占める領域の座標を 1 として、空白部分の座標を 0 として 1 である部分の x、y 座標を線形回帰で行い、直線の傾きが求められる。傾きをオイラーアングルに変換して、最短外径を把持する角度がわかる。物体の中心点を通過する最短外径を計算し、物体の把持姿勢とする。ロボットアームに把持対象の位置情報と把持姿勢情報を送信し、最適な把持動作を行う。

4. 片付けロボットのハードウェア

実験には、片付け対象として、卓上にペットボトル、紙コップとボール 3 種類の物体を複数放置した。ロボット Sciurus17 は起動した後、卓上を見て、物体検出を行いながら、ユーザの指示を待機する。YOLACT による把持実験は、学習された名前をもとに移動する場所を指定して、指定先に物体を移動することができれば成功、実験範囲は、RealSense D415 の検出範囲内を対象として、ランダムに物体を配置する。

まず、YOLACT による物体認識を行い、YOLACT による物体認識結果を Fig.3 に示す。次に、「ペットボトルを片付けて」という音声指示を理解し、抽出したペットボトルは検出した対象とマッチングでき、それから、片付け対象として確定する。ボールの中心座標と最短外径を取得し、把持姿勢を確定し、Sciurus17 のロボットアームに送信することで、対象物体のペットボトルが事前に指定される場所に片付けられる。物体把持動作の流れは Fig.4 に示す。Fig.4 (a)に示す初期ポーズから片付け対象を確定し、最適把持姿勢で対象物体を把持する(Fig.4 (b)に示す)。物体を把持されてから持ち上げ操作、運搬操作と放置操作(Fig.4 (c)に示す)を通じて、事前に指定した場所に片付けられる。



Fig.3 YOLACT による物体検出結果

物体の数、種類、位置を変更し、計 15 回実験を行った。15 回の実験結果から、15 回の実験結果から成功率を算出してまとめたものを Fig.5 に示す。把持の成功率が計 95% 以上である。ただし、音声命令が認識できない場合はよく出てくる。ペットボトルの物体認識率が 90% 以上だが、把持成功率が 86.7% になり、比較的が悪かった。原因としては 3D 点群データから線形回帰を行うときに、ノイズが入る原因で把持座標の算出する際に、誤差が生じて、線形回帰の結果に、傾きが実際とずれがあると挙げられる。また、音声命令の誤認識原因としてはまずノイズが入って来て、認識の正確率に影響を与える。また、人による発音が少し

違うと考えて、発音の影響で認識率にも下がるということが分かる。そこで、問題を解決するために、音声認識の時、正しい結果を認識できるまでに、何回も聞き取りを繰り返すと設定した。



(a)初期ポーズ (b)物体把持 (c)物体運搬

Fig.4 物体把持動作

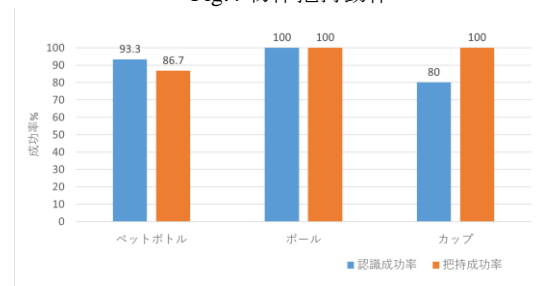


Fig.5 実験結果

5. おわりに

本研究では、音声指示を対応した片付けロボットの制御システムを提案した。音声認識による把持対象を確定し、YOLACT を使い物体を認識し、カテゴリを行った。把持姿勢は物体の最短外径と中心座標で確定する。実機実験により、物体の把持座標を取得し、指定した場所に物体を片付けることができた。今後は、より複雑な環境を考慮して、立っている物体をどのように認識し、物体が立っている状態と倒れている状態でも区別できて、かつ物体を正しく認識し、物体の把持姿勢を生成するが望まれる。物体認識方法は、複数のカメラを用いて、違う角度から物体を観察し、得た情報を融合するという手法が可能になると思う。また、物体認識の上、物体の材質、中身の重さにも判断できて、物体を把持する最適な力と物体を持ち上げるため、最適な力を計算し、最適な把持動作が行われる研究が必要だと考えられる。また、未知物体に対して、どのように対応する方法について考える必要がある。

参考文献

- [1] 中村 友昭, 長井 隆行, 岩橋 直人, “ロボットによる物体のマルチモーダルカテゴリゼーション”, 電子情報通信学会論文誌, D, 情報・システム 91(10), pp.2507-2518 (2008).
- [2] G. LYU, H. YIN, X. YU and S. LUO: "A Local Characteristic Image Restoration Based on Convolutional Neural Network", IEICE TRANS. INF. & SYST., VOL.E99-D, NO.8 (2016).
- [3] J. Han and K.K. Ma: "Fuzzy Color Histogram and Its Use in Color Image Retrieval", IEEE TRANSACTION ON IMAGE PROCESSING, VOL.11, NO.8, pp.944-952 (2002).
- [4] J. Redmon and A. Farhadi: "YOLOv3: An Incremental Improvement", arXiv : 1804.02767 (2018).
- [5] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee, "Yolact: Real-time instance segmentation." Proceedings of the IEEE/CVF international conference on computer vision, pp.9157-9166 (2019).