

ロボットのための動的観察に基づく物体認識手法の提案

三浦 且之 福田 悠人 小林 貴訓 久野 義徳
埼玉大学

1. はじめに

少子高齢化の進行に伴い、サービスロボットの需要が高まっている。我々は、人が指示した物体を代わりに取ってきてくれるロボットの研究、開発を行っている。このようなロボットの実現には周囲の物体を認識する技術が不可欠である。近年では、深層学習を用いた画像物体認識が高い認識性能を実現しているが、実環境では誤認識も多い。深層学習では通常ネット上にある画像を学習データに用いており、それらの画像には、コップの取っ手のような、その物体らしさが分かる特徴が顕著な角度から撮影された画像(以下「ファインショット」と呼ぶ)が多い。このようなファインショットの画像を入力画像として認識を行うと、高い認識性能を確認できる。一方で、実環境ではファインショットを得られることは少ないことが原因だと考えられる。

ロボットは自ら移動できるので、物体の周辺を移動させて、複数視点から物体を観察することで認識に有用なファインショットを自ら探し出すことができる。そこで本稿では、ロボットの動的観察によりファインショットを得ることで正しく物体認識を行えるかを検討する。

2. ロボットの動的観察による物体認識手法

本稿では、移動ロボットに SoftBank Robotics 社の NAO を用いる。また、物体認識には深層学習で高い認識性能をほこる GoogLeNet[1]を用いた。ファインショットを見つけるために、物体を中心にロボットを円形に一周させ、その円周上の 8 箇所においてロボットのカメラから物体画像を撮影する。各視点で得られた画像に対し逐次物体認識を適用し、得られた認識結果の上位 2 位までの候補を確信度と共に取り出す。その認識結果からファインショット判定を行いファインショットと判定されれば、その時点の認識結果を出力して探索を終了する(図 1)。ファインショットは、1 位の確信度が 80%以上、または 1 位と 2 位の確信度の差が 60%以上のものとする(図 2)。8 視点の全てでファインショットを得られなかった場合は、それぞれの認識結果から最も確信度が高いものを認識結果として出力する。

3. 実験

提案手法の評価のため実験を行った。20 個の物体について、本手法により認識を行い、また比較として 8 視点全てで撮影した画像に物体認識を適用した。



図 1. 動的観察に基づく物体認識



図 2. ファインショットの例

本手法による結果として、ファインショットが得られたものは 11 個であり、それらは全て正しく認識できた。またファインショットが得られなかったものは 9 個あり、そのうち正しく認識できたものは 6 個、誤って認識されたものは 3 個であった。ファインショットが得られない原因として、ロボットの視点移動が円周方向のみであり、仰角方向に視点移動ができなかったことが考えられる。

一方、ロボットが撮影した 8 視点全ての画像について物体認識を適用した結果、正しく認識した画像は 160 枚中 87 枚で認識率は約 54%であった。提案手法では、20 物体中 17 物体を正しく認識でき認識率が 85%であることから、ロボットの物体認識における動的観察による有効性を確認できた。

4. おわりに

本稿では、ロボットの動的観察によりファインショットを探し出すことで、物体を正確に認識する手法を提案した。実験を行い、ファインショットの物体認識に対する有効性と多視点から得た画像を元に認識を行うことの有効性を確認した。

今後は、ロボットが物体に応じて視点移動を戦略的に行うようなシステムの導入を検討する。また、ロボットの視点移動による物体の距離や見えの変化に基づき、物体形状やサイズ情報を得て認識を行う手法について検討する予定である。

本研究の一部は JSPS 科研費 26240038 の助成による。

参考文献

- [1] C. Szegedy et al., "Going Deeper with Convolutions," *Proc. of CVPR*, 2015.