

複数視点間の認識結果の統合に基づく物体認識

渡邊 真悠[†] 福田 悠人^{††} 小林 貴訓^{††} 久野 義徳^{††}

[†] 埼玉大学工学部情報システム工学科 ^{††} 埼玉大学大学院理工学研究科

1. はじめに

近年、少子高齢化社会の進行による介護者の人手不足が深刻な問題となっている。その中で、介護の分野におけるロボットの活躍が期待されている。我々は、人の代わりに指示した物体を取ってきてくれるサービスロボットの研究・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物体認識の技術が必要不可欠である。近年では深層学習の技術を画像認識の枠組みに適用して物体の認識を行う手法が開発され、高い認識精度を実現している。しかし、どのような状況においても誤認識せずに物体を認識することは困難である。特に実際の環境では、以下のような要因が画像からの物体認識をさらに困難にすると考えられる。

まず、撮影状況によって、画像中の物体の特徴が正しいクラスの特徴よりも別の物体のクラスの特徴と一致してしまう場合である。もう1つはオクルージョンや背景の写りこみにより対象物体の特徴を上手く抽出できない場合である。これらの問題は画像を取得する環境及び物体に対する視点位置に起因するものである。本稿ではこの問題を解決するために複数視点の画像を用いて物体認識を行い、それらの結果を統合することで、認識精度を向上させる手法を提案する。

2. 認識結果の統合手法

実験データとして、テーブルの上に異なる複数の物体(4~5個)を適当に配置し、それらの配置を変えないまま任意の異なる視点から全ての物体が映るように画像を撮影した。物体のセグメンテーション(領域分割)は正しく行われているものとし、それぞれの領域に対し物体認識を適用させた。本稿では、物体認識にCNN(Convolutional Neural Network)に基づいて認識を行うCaffe[1]を用いた。Caffeにより得られた認識結果の上位5位の候補をスコア(その認識結果の確からしさを示す値)と共に取り出す。

認識結果の統合については、視点間の物体の対応付けが不可欠である。一般的には各カメラ視点の位置を事前に校正し、各画像間の物体領域の対応付けが行われるが、本研究ではこのような制約を与えず、全ての組み合わせに対し、同一のラベルが存在すればスコアを足し算で合算し、スコアの統合を行う(図1)。その結果として、最も高いスコアを出力した組み合わせを最適な組み合わせとして、最終的な物体の認識結果として出力する。

3. 実験

実験データとして19組の画像セットを用いた。結果として、9組のデータセットにおいて認識結果

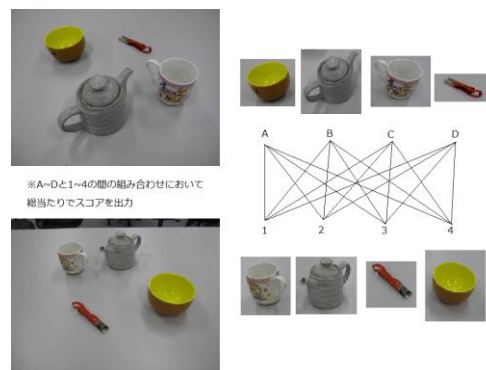


図1. 統合手法の概要

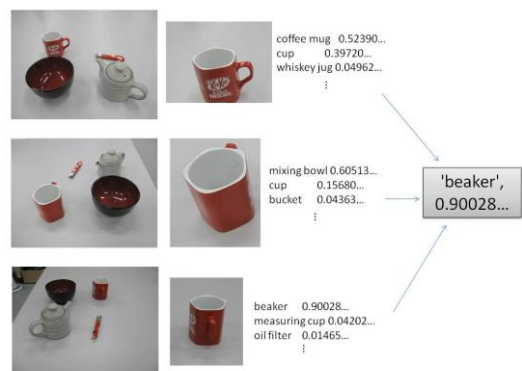


図2. 統合により劣化した場合

が改善された。また他の5組はいずれかの視点における認識結果を維持し、3組は劣化した。別の2組は最も高いスコアを出力する組み合わせは一意に得られず、物体認識結果の改善も見られなかった(図2)。

4. おわりに

複数視点間で同一のラベルが候補として出力されなかったことに起因して、組み合わせが一意に得られなかった場合があった。これは、一視点における認識結果から得られる候補を増やすことで改善されると考えられる。さらに、認識結果の統合により結果が劣化してしまった原因は、一方向の画像から特徴が十分に抽出されないことが原因であると考えられる。これらの問題を解決するため、今後は、カメラの視点位置に関する制約を導入することを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] J. Yangqing et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," *Proc. of CVPR*, 2014.