

インタラクションに着目した物体認識

石島 樹[†], 石井 雅樹[†], 下井 信浩[†]

秋田県立大学[†]

1. はじめに

近年, 人間の生活環境で共存することを目的としたロボットの開発が急速に進んでいる. ロボットが人間と共存するためには高度な認識機能を有することが必要不可欠である. 物体認識や環境認識を行う際, 人間は視覚から多くの情報を獲得し, その情報の関連性や自身の経験を基に総合的に判断している. すなわち, 人間の認識機能には, 「人」が「環境」の中で「物体」をどのように扱っているかといったそれぞれの相互関係(インタラクション)が大きく寄与していると考えられる.

そこで本研究では, 人と物体とのインタラクションに着目した物体認識手法を提案する. 本稿では基礎検討として, 物体に対する「掴む」という動作に着目し, 獲得したシーン画像に対し, 一般公開されている WebAPI[1]を活用した物体認識を試みたので報告する.

2. 研究概要

2.1 提案手法

本研究では, 物体と人間とのかかわり方を物体の属性情報として抽出し, その属性情報を基に物体認識を実現することを目標としている. 本研究では物体の属性情報を動的属性と静的属性として定義する. 前者は人間が物体に対してどのように関与しているかを表す動詞であり, 後者は従来のエッジや色などの幾何学特徴を用いた物体に関する情報である.

人間が物体を扱う場合, 多くは手で触れると予想される(持つ, 掴む等). 本研究では動的属性情報として手の動きに着目した. 具体的には固定カメラ(ELECOM社製 UCAM-DLK130TRD)を用いて物体を扱う際の動画像を撮影し, 指を検出[2]することにより「掴む」動作を動的属性として抽出した. 次に掴んだ瞬間の静止画像を取得し, その画像を WebAPIに入力することにより物体認識を行った.

2.2 WebAPI

API は外部から利用したい機能を容易に取り入れることができるため, 開発コストを軽減できる. また, ネットワークを介したリアルタイムでの利用が可能である. 本研究では AlchemyAPI 社が公開している AlchemyVision の Image Tagging API を用いた. Image Tagging API は画像解析系のサービスであり, DeepLearning(深層学習)が応用されている.

無償利用に際しては, 画像サイズが 1MB 以下, 回

Object recognition focused on interaction with human being

[†]Tatsuki Ishijima, Masaki Ishii : Akita Prefectural University

数は 1 日あたり 1000 トランザクション, 同時リクエストは 5 件まで可能である. API に認識対象となる画像名や画像の URL を与えることにより物体名(タグ名)と適合率(スコア)を得ることができる.

2.3 指認識

「掴む」動作を動的属性として抽出するため, 以下の手順で指認識を行った. 図 1 にそれぞれの処理結果を示す.

- ① HSV 色空間を用いた肌色抽出
- ② ノイズ除去のためのフィルタ処理
- ③ 肌色領域の輪郭抽出
- ④ 輪郭座標を一定間隔で取得
- ⑤ 隣接座標から外積, 内積を算出
- ⑥ 外積, 内積の閾値判定による指認識

指を認識した後, 指の本数や隣接する指の距離を用いて, 人間が物体を「掴む」動作を検出した. 図 2 に物体を掴む瞬間に取得した画像を示す.

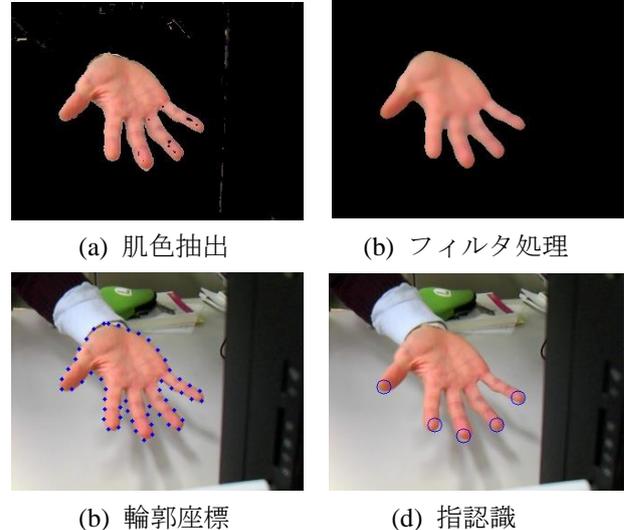
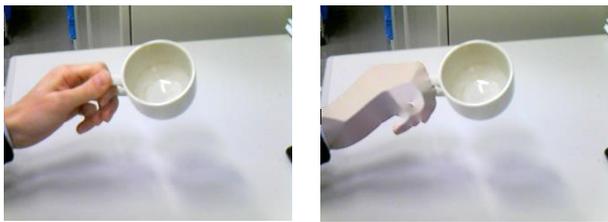


図 1 指認識アルゴリズム



図 2 掴む瞬間の画像



(a) 原画像 (b) 手を除去した画像



(c) 手の除去, トリミングを施した画像
図3 使用画像の一例

3. API を用いた物体認識

物体を掴んだ時の画像を対象とし, API を用いた物体認識実験を行った. 本稿では対象物体をマグカップに絞り, 様々な角度から取得した 30 枚の画像を使用した (①). また, ①の画像から手を除去した画像 (②), および, ②の画像からマグカップを中心にトリミングした画像 (③) を取得し, それぞれの認識結果を比較した. なお, API の認識結果において物体名が獲得できない場合は認識不可とした. 実験に用いた画像の一例を図3に示す.

図4, 図5, 図6に認識結果を示す. 結果として物体名が得られた画像の多くは, カップの内側が写った画像であった. また, 物体名の多くはカップの内側のもの(コーヒーなど)を予測しているような結果が得られた. 図4および図5を比較すると, 前者の認識結果には person が含まれているものの, 後者の結果には含まれていないことがわかる. すなわち, 手の入り込みが認識結果に大きく影響していると考ええる. また, 図6の認識結果では飲み物名他に容器名が得られている. これはマグカップ以外の手や背景といった不要な領域を除いたためと考える.

4. 結論

API を用いた物体認識では認識させたい物体のみを抽出し, 背景などの物体以外の情報を除去することが有効である. また, マグカップを対象とした実験ではシーンを認識しているような結果が得られたことから「掴む」という動作を組み合わせることで物体名の絞り込みが可能と推測される.

今後は API と指認識を組み合わせた物体認識手法および動的属性情報を用いた物体認識手法について検討を加える予定である.

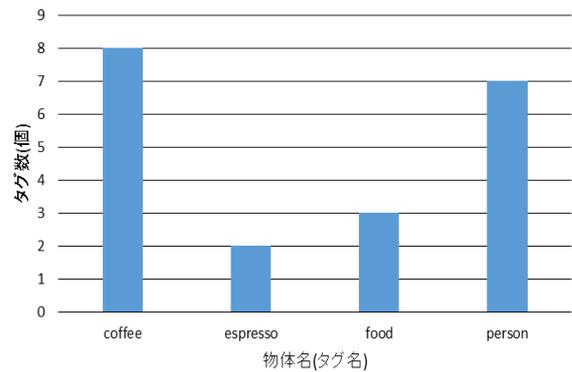


図4 原画像を対象とした認識結果 (①)

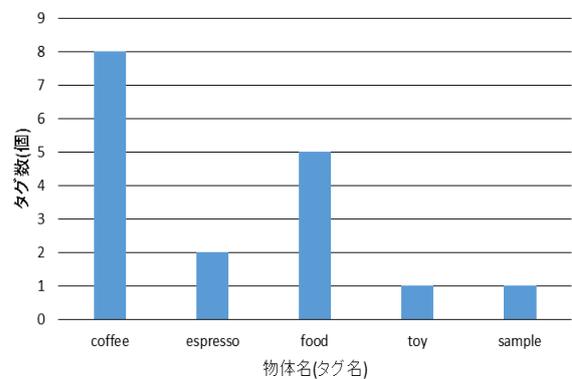


図5 手を除去した画像を対象とした認識結果 (②)

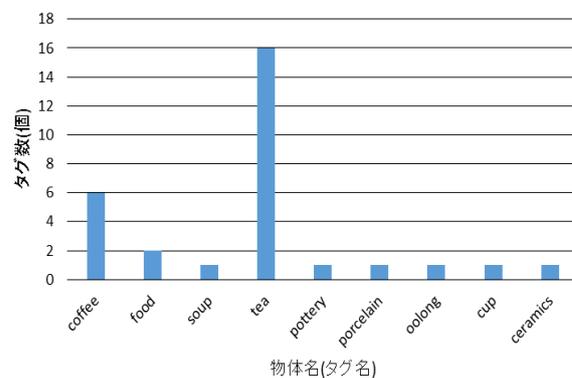


図6 手の除去, トリミングを施した画像を対象とした認識結果 (③)

参考文献

- [1] AlchemyAPI Powering the New AI Economy <<http://www.alchemyapi.com/>> (2015.1)
- [2] 藤本美加, 井上聡: 「手話認識のための手形状認識アルゴリズム」, 人工知能学会全国大会(第28回)論文集, p63 (2014)