

物体の機能に基づく認識

田中 雄翔^{1,a)} 滝口 哲也^{2,b)} 有木 康雄^{2,c)}

1. はじめに

一般物体認識 [3] はコンピュータビジョンにおいて一つの大きな分野となっている。このために数多くの研究が行われ、近年では以下のパイプラインが標準となっている。

(1) 訓練サンプルにおける局所記述子 (SIFT[2] など) を抽出。(2) 各画像の局所記述子から、それぞれの画像を表す特徴量 (BoW 表現 [1] など) を抽出。(3) 各画像の局所記述子を統合した特徴量を用いて、画像を分類 (サポートベクターマシン (SVM) など)。

実用化に耐えうる物体認識の精度が保証されれば、画像中のどこに何の物体が映っているのかがわかるようになるため、たとえばユーザーが、“A という物体はどこにあるか。”という質問をロボットにするだけで、ロボットが自動で環境内を探索し、目的の物体を持ってくることなどへの活用が期待されている。

しかし、そういった場面において、ユーザーが本当に求めているものは、A と呼ばれる物体なのではなく、自分の要求を満たす機能を備えたもの、であると考えられる。具体的に説明するために、ユーザーが、“コップを取ってください。”とロボットに命令した状況を考える。このときユーザーがコップと呼ばれる物体を要求した背景には、喉が渴いたから水が飲みたい、といったユーザーの欲求が潜んでいると考えられる。その上でユーザーは、水が飲みたいときにはコップに水を注ぎ入れることで簡単に水が飲める、ということを知っているため、水を容れるという機能を備えた物体であるコップを指定する。つまり、ユーザーが本当に求めているものは、コップと呼ばれる物体なのではなく、“水を容れることができる”物体と考えられる。

このことと従来の物体認識との違いは、“コップという名称で呼ばれていない物体でも、ユーザーの要求を満たす可能性がある”，ということである。極端に言えば、「コップを取ってください。」というユーザーの命令に対してロ

ボットは、窪みがあり耐水性がある物体であればユーザーの要求を満たすと推定し、ユーザーに提供してもよい。

つまり、従来の物体認識で用いられてきた画像パターンを表現する特徴 (SIFT など) ではなく、物体の構造や素材、物理量といった従来用いられてこなかった特徴を用いて、物体の機能がどういった特徴から発現しているのか、を学習する必要がある。

こういった、機能に関する物体認識ができるようになれば、たとえば非常時でまわりにコップがない状況において、ユーザーが水を容れることができるものはコップである、という知識しか持っていない状況でも、ロボットが、コップと呼ばれるものでなくともユーザーの要求を満たすことができる物体を探すことができるようになり、さまざまな場面での活躍が期待できる。

こういった背景から本研究では、物体の画像パターン認識ではなく、物体の機能に基づいた認識について研究を行うと考えている。

2. アプローチ

前述したような、実世界で物体の機能を推定するアプローチには、図 1 に表すように 2 通りの状況があると考えられる。

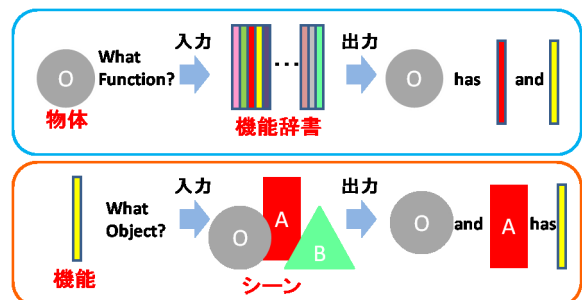


図 1 実世界での推定アプローチ

ここで、システムが持っている機能に関する知識（その機能を発現するために不可欠な要素などをチェックリストなどの形で記したもの）を機能辞書と定義する。

(1) ボトムアップアプローチ (図 1 上)

ある物体に対して、その物体が機能辞書の中のどの機能を持つのか、を推定する。

¹ 神戸大学システム情報学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

² 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

a) ytanaka@me.cs.scitec.kobe-u.ac.jp

b) takigu@kobe-u.ac.jp

c) ariki@kobe-u.ac.jp

(2) トップダウンアプローチ (図 1 下)

ある機能に対して、どの物体がその機能を発現することができるのか、を推定する。

以上の 2 通りのアプローチについて、実用的なシーンでの活用について考察する。(1) のボトムアップアプローチは、ユーザーがある物体に対して、システムに「この物体は何に使えますか。」と問う場合である。つまり、ユーザーが知らない、もしくは思い付かないような機能をシステムが推定し、返答しなければならないため、非常に難しい問題であるといえる。

これに対して (2) のトップダウンアプローチは、ユーザーの要求に応じた機能を持つ物体を、環境内から探すという問題であるため、実世界での活用においても有用であり、こういった素性からその機能が発現しているのかを考えようとする本研究では、理にかなった問題設定であるといえる。

以上のことから本研究では、(2) のユーザーの要求を満たす機能を持つ物体を推定するトップダウンアプローチに着目する。

2.1 機能辞書

機能辞書とは、ある機能について、その機能を発現するために必要な素性を表したものである。たとえば、椅子が「座ることができる」という機能を持つには重量への耐久性、座る部分の水平性などを考慮する必要があり、人間は椅子に座る前には、自分が持つ機能辞書に基づいてその椅子が本当に「座ることができる」物なのかを推定してから実際に座っていると考えられる。

具体的にどういった素性がその機能に関係するかはそれぞれの機能によっても異なると考えられるため、得られる情報から必要な情報を取り出す特徴選択の考え方も必要になると思われる。

2.2 物体の構造

機能がどういった素性から発現しているのかは未知数であると述べたが、機能を学習するためには、

(1) 実際に人と物がインタラクションしているシーンをベースにする

例：字を書くことができる、インターネットにアクセスできる。

(2) 物の構造をベースにする

例：座ることができる、水を容れることができる。

という 2 つの場合が考えられる。

(1) の実際に人と物がインタラクションしているシーンをベースにする方法は、ボールペンやパソコンなど、複雑な内部構造やシステムを持つ人工物によって発現するものが多いと考えられ、難しい問題である。

そこで本研究では、(2) の場合を想定し、システムに搭

載されたカメラで撮られた映像から、物体の三次元復元を行い、物体の構造やそれぞれのパーツと機能との関連性について考察していくというアプローチを取る予定である。

3. 実装予定

画像から物体の三次元復元を行う手法は数多く提案されている [4][5]。しかし、復元されたモデルの内部構造やパーツまで考慮している研究は少ない。

たとえば椅子であれば、腕かけ、背もたれ、キャスターなどのようにパーツ分けすることができるが、ユーザーの座りたいという要求を満たすためには、実際に座る部分さえあればよい。

このことから今後は、機能と物体の数を具体的に想定し、図 2 のように、物体に関して三次元計測を行った後、さらにその物体の基本的なパーツを推定する。その後、特徴量を抽出し学習することで、どの特徴がその機能を発現しているのかを考察する予定である。

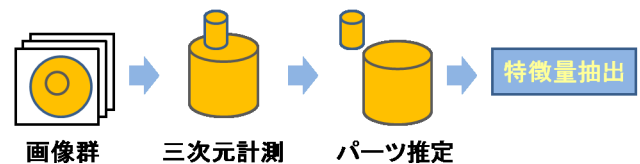


図 2 実装予定フロー図

4. まとめ

本研究では、物体の機能に基づく認識について考察し、画像パターンに基づく従来の物体認識との違いや、考えられるアプローチ、物体の機能に基づく物体認識の有用性について論議した。

今後は具体的な実験を通して、機能というものがどういった素性から発現しているのかを、考察していく予定である。

参考文献

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of key-points," *Proc. ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1-22, 2004.
- [2] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. IEEE International Conference on Computer Vision*, pp. 1150-1157, 1999.
- [3] 柳井 啓司, "一般物体認識の現状と今後," *情報処理学会論文誌:コンピュータビジョンとイメージメディア*, Vol. 48 No. SIG 16(CVIM 19), 2007.
- [4] Jan J. Koenderink, Andrea J. Van Doorn, "Affine structure from motion," *Opt. Soc. AM. A*, Vol. 8, No. 2/February 1991.
- [5] Sudipta N. Sinha, Drew Steedly, and Richard Szelisk, "A multi-stage linear approach to structure from motion," *Trends and Topics in Computer Vision*, Springer, 2012.