

自己組織化マップを用いた物体概念の学習

Learning of object concept using a self-organizing map

藤本智也^{*1}
Tomoya Fujimoto

田口亮^{*1}
Ryo Taguchi

服部公央亮^{*2}
Koosuke Hattori

保黒政大^{*2}
Masahiro Hoguro

梅崎太造^{*1}
Taizo Umezaki

^{*1} 名古屋工業大学
Nagoya Institute of Technology

^{*2} 中部大学
Chubu University

Household robots need a mechanism to learn names of objects through interaction with users in daily life. When users teach a robot objects' names, viewpoints of objects are unknown. Therefore a robust mechanism to changes of viewpoints is necessary in learning system of objects' name. In this paper, we propose a mechanism to learn objects' names by using self-organized Map(SOM) and SVM. SOM classifies the viewpoints of images for each objects without teacher data. SVM learns an object of a single classified viewpoint.

1. はじめに

近年、日常社会の中で人と密接に関わりながら活動するサービスロボットの研究・開発が進んでいる。しかし、実世界の物体の認識を含むタスクを行うためには、物体の名称と特徴の対応関係(物体概念)を人とロボットの間での共有が必要である。

中村らは、複数モダリティの特徴を用いた物体概念の獲得手法を提案している[中村 2008]。この研究では、視覚特徴を抽出する際に、物体をターンテーブルに乗せて全周画像を撮影した。しかし、実運用の場面では、ロボットの目の前で人が物体を持ち、任意に回転させながら物体の名称を教えることが想定される。その場合、物体の姿勢は未知となり、さらに物体の動かし方に依存して得られる視覚特徴が不均一となるため識別性能が低下する。そこで本研究では、自己組織化マップ(SOM)[Kohonen 1995]を用いることで、物体の姿勢変化にロバストな物体概念学習手法を提案する。

2. 使用したロボットと教示状況

本研究では図 1 に示す、移動ユニットに RGB-D カメラを乗せたロボットを実験に用いた。図 1 のように教示者がロボットに物体を見せ、物体を任意に移動、回転させながら、その名前を教示する。本実験では名前の教示はテキストで与えた。

実験環境は、サービスロボットが活動するであろう、人が生活している空間(研究室内)を設定した。ロボットの周囲には教示者、教示対象物体以外にも、数多くの物体が存在している。そのため、ロボットのカメラには背景として別の物体も写りこむ。

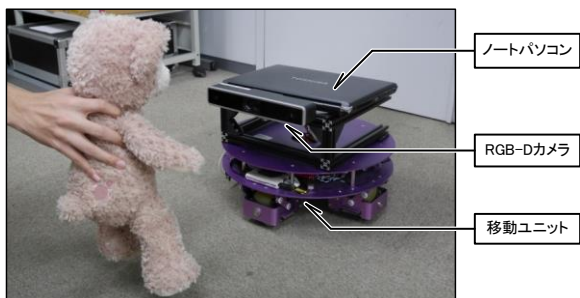


図 1 ロボットの構造と教示状況

3. 提案手法

3.1 物体概念の学習

物体概念の学習機構は、局所特徴量として SIFT、特徴量の表現に Bag of Features(BoF)、特徴群のクラスタリングに SOM、識別器に SVM を用いて設計した。

提案手法で学習する際の処理の流れを以下に示す。

Step1. 特徴抽出

- (1)画像から SIFT 特徴を抽出
- (2)画像中の SIFT 特徴をヒストグラムで表現(BoF)

Step2. 特徴学習

- (1)SOM によりヒストグラムをクラスタリング
- (2)SOM のユニットごとに SVM を学習

3.2 RGB 画像, 深度画像からの物体特徴抽出

RGB 画像, 深度画像から物体の特徴を抽出する。まず深度画像のエッジ情報から複数の領域に分ける。最も近い領域を教示物体と判定し、図 2(b)に示すような物体のマスク画像を作成する。物体のマスク画像が物体の輪郭を表していることに注目し、マスク画像から SIFT 特徴量を抽出する。この特徴を物体の輪郭特徴と定義する。また物体のテキストチャからも特徴を得るため、RGB 画像全体から SIFT 特徴を抽出し、その後、マスク画像を用いて背景に含まれる特徴点を除去する。



(a)深度画像 (b)物体のマスク画像 (c)テキストチャ切り出し
図 2 マスク処理結果

3.3 特徴の表現法

(BoF)を用いて特徴を表現する。室内環境下で無作為に撮影した画像から、ベクトル量子化のための code book を作成した。提案手法では、マスク画像から抽出できる特徴点数が、テキストチャから抽出できる特徴点数に比べて少ないことから、物体の輪郭特徴を 50 次元、テキストチャの特徴を 500 次元と定めた。そして、一組のカメラ画像と、深度画像から抽出した特徴を、550 次元のヒストグラムとして表現する。輪郭特徴とテキストチャ特徴の

連絡先: 藤本智也, 名古屋工業大学工学部,
〒466-8555 名古屋市昭和区御器所町, TEL:052-735-7450,
fujimoto@ume.mta.nitech.ac.jp

code book の次元数が異なるため、比率を合わせるために輪郭特徴は和が 0.1 にテクスチャ特徴は 1 になるように正規化する。この処理により、画像から抽出した多数の特徴点を、一つのヒストグラムとして表現可能となる。そして、それぞれの画像から抽出した特徴点の数が異なる場合も、同じ枠組みで比較できる。

3.4 特徴の学習

各物体の画像は、様々な姿勢の物体を人が任意に動かしながら撮影した画像であり、角度や姿勢がそれぞれ異なる。この時、平面上の回転変化やスケールの変化は SIFT により吸収され、同じ特徴が抽出されるが、SIFT で吸収できない変化(ロール, チルト)をした画像については、同じ物体でも異なる特徴が抽出される。そこで、同一物体の特徴でも、それぞれ似た特徴ごとに、分類してから学習することが必要となる。そこで、提案手法では、入力データの分類に自己組織化マップ(SOM)を用いた。SOM はデータを教師なしでクラスタリングできる。1つの物体につき 1 つの SOM を用いることで、同一物体内での特徴変動、すなわち物体の角度の違いに応じてデータが分類できる。SOM の類似度計算には Bhattacharyya 係数を用いた。全物体の SOM を学習した後、各ユニットに分類されたデータ群をそれぞれ別のクラスとして one-versus-one によるマルチクラス SVM を学習する。識別には SOM は用いず、SVM のみを用いる。

3.5 実験条件

家庭などでロボットが目にする対象として、図 3 に示すように玩具や本など 7 個の物体を用意した。物体を人が手で動かしながら、0.1 秒ごとに 10 秒間、計 100 枚の画像を撮影する。背景の異なる 3 ヶ所で行うことで、1 つの物体あたり 300 枚の画像を撮影した。2 ヶ所で得られた 200 枚を学習データ、残り 1 ヶ所で得られた 100 枚を評価データとして、交差確認法で評価する。SOM はランダムに与えられる初期値の影響を受けるため、複数回試行した。また、SOM のサイズは 2×2 とした。



図 2 実験に使用した物体

3.6 実験結果と考察

実験結果を図 3, 図 4 に示す。比較のため、SOM を用いない SVM 単体の結果および、輪郭情報を使用しない結果も併せて示す。SVM のみを使用した場合と、提案手法を用いた結果を比較すると識別性能が向上していることがわかる。この要因として、同一物体の画像でも抽出できる特徴は異なるため、類似しているデータごとに分けることにより、特徴のパラつきが少なくなり、適切な学習が行われたことが考えられる。また、深度画像から物体領域を切り出して作成した、マスク画像の輪郭情報を、特徴として用いることで、識別性能が向上した。

さらに表 1, 表 2 に提案手法を使用した時の混同行列を示す。縦軸に入力した画像の真のラベル, 横軸に識別器が出力したラベルを示している。テクスチャ情報のみでは、(d)イチョウや(f)ペットボトル, (g)ネコのぬいぐるみの識別率が低い。この要因と

して、テクスチャの変化に乏しい部分が多いことやテクスチャ内のエッジが少ないことが挙げられる。一方で、表 2 に示すように、テクスチャ情報と輪郭情報を併用した場合には、それらの物体の識別率が向上し、全体としての識別率が上昇している。このことから物体の輪郭情報を特徴に用いることにより、物体の特徴を効果的に捉えることが出来るかと結論付けられる。

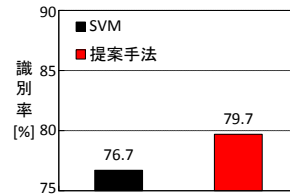


図 3 テクスチャ情報のみ

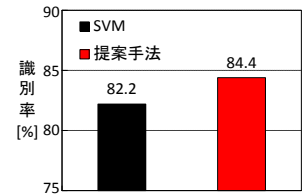


図 4 テクスチャ+輪郭情報

表 1 提案手法(テクスチャ情報のみ)の混同行列 [%]

入力 \ 出力	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a)	88.3	1.0	0.3	0.0	2.7	5.7	2.0
(b)	0.3	87.7	3.0	1.3	1.7	3.7	2.3
(c)	0.3	3.3	86.0	2.0	4.0	1.3	3.0
(d)	0.7	2.0	6.0	75.0	7.3	2.7	6.3
(e)	1.0	2.3	2.0	1.7	89.0	1.3	2.7
(f)	9.0	3.3	2.3	7.0	4.7	66.7	7.0
(g)	5.3	1.7	5.3	8.3	7.7	6.3	65.3

表 2 提案手法(テクスチャ+輪郭情報)の混同行列 [%]

入力 \ 出力	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a)	88.7	0.3	0.3	0.3	1.7	5.7	3.0
(b)	1.3	85.7	2.3	1.3	1.3	6.7	1.3
(c)	0.3	4.0	85.7	0.7	3.7	2.3	3.3
(d)	2.0	1.3	3.7	84.3	1.7	3.0	4.0
(e)	0.7	2.0	2.7	1.0	90.0	1.3	2.3
(f)	5.7	2.7	3.7	3.3	3.0	77.3	4.3
(g)	4.3	1.7	2.3	5.7	3.3	3.3	79.3

4. まとめと今後の課題

本研究では、ロボットが人とのコミュニケーションを通じて、学習を行う際に必要な、物体概念学習手法を提案した。同一物体でも姿勢の違いにより、得られる特徴量が変化するという問題に対して、自己組織化マップを用いて対応した。物体の識別精度の向上が見られたほか、テクスチャだけでなく、物体の輪郭情報も学習することにより、テクスチャから抽出出来る特徴が乏しく、テクスチャの情報だけでは識別率が低い物体に対しても一定の識別性能を得ることに成功した。

今回用いた SIFT は色情報を使用していないため、今後の課題として、色情報も特徴として捉えるなど手法の改善による識別性能の向上が挙げられる。また、実ロボットでのリアルタイム物体学習を行うために、音声対話システムと統合し、人とのコミュニケーションを通じた物体概念学習の実現を目指す。

参考文献

[中村 2008] 中村友昭: 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, 電気情報通信学会論文誌, Vol.J91-D No.10, pp.2507-2518, 2008.
 [Kohonen 1995] Kohonen.T: Self-Organizing Maps, Springer Series in Information Sciences, Volume 30, 1995.