論文

物体形状を考慮した dense CRF による機能属性認識の高精度化

非会員 飯塚 正樹*a) 正員 秋月 秀一*,** 正員 橋本 学*

Accuracy Improvement of Functional Attribute Recognition by Dense CRF Considering Object Shape

Masaki Iizuka*a), Non-member, Shuichi Akizuki*,**, Member, Manabu Hashimoto*, Member

(2017年12月18日受付, 2018年5月16日再受付)

In this paper, we propose a method to recognize functional attributes of everyday objects for vision system of partner robots. On the related research, there is a method to optimize recognition result with dense (fully connected) CRF which use the estimation result of functional attribute for each pixels. However, since this method is optimized from RGB data, it isn't able to sufficiently consider the shape of object, which have a relationship with the function attribute. In the proposed method, the recognition accuracy of functional attributes is improved by considering the object shape with the dense CRF describing the three - dimensional positional relationship. As a result of the experiment, the recognition rate of the proposed method is 77.0 %, which is 3.8 % higher than the related method. In addition, we confirmed that the processing speed is high as a side effect by reducing processing cost by oversegmentation of input data and using high speed identification by Random Forests. The mean processing speed per an object was 109ms in the proposed method.

キーワード:機能属性, アフォーダンス, 生活支援ロボット, dense CRF **Keywords:** functional attribute, affordance, partner robot, dense CRF

1. はじめに

近年,公共施設や家庭内における利用を想定した生活支援ロボットの実用化に対する期待が高まっているが,ここには生産システムの自動化を目的とした知能ロボットとは異なる課題がある。最も大きな差違としては,製造現場ではロボット操作対象物に関してCADモデルのようななんらかのモデル利用を前提とした特定物体認識が主課題になるのに対して,家庭環境下では一般にモデル利用が困難な一般物体認識が求められることが挙げられる。例えば,ユーザがロボットに「コップにジュースを注いで持ってきて」と指示するような単純なタスクであっても,2つの課題,す

a) Correspondence to: Masaki Iizuka. E-mail: iizuka@isl.sist. chukyo-u.ac.jp

 * 中京大学大学院 工学研究科 〒 466-8666 名古屋市昭和区八事本町 101-2 Graduate School of Engineering, Chukyo University 101-2, Yagoto Honmachi, Showa-ku, Nagoya 466-8666, Japan
 ** 慶應義塾大学 〒 223-8522 横浜市港北区日吉 3-14-1 Keio University 3-14-1, Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan なわち個々の対象物に関する形状や見えのモデルを利用す ることなく対象物の品種(クラス名)を識別し(課題1), さらにそれに対する適切なロボットオペレーションを自動 生成しなければならない(課題2)。

第1の課題はいわゆる一般物体認識の問題であり, YOLO⁽¹⁾やSSD⁽²⁾に代表される Deep Learning ベースの 手法によって解決されつつある。しかし,第2の課題を解 決するためには,対象物のクラス名を推定した後に,コッ プ内のジュースを注ぐ位置や,コップを把持するための位 置を推定する必要がある。すなわち,初めて見る対象物に ついて,"液体を蓄える",あるいは"(手で)把持される" という「役割り」を担う領域を認識する必要がある。この ような,対象物が担う「役割り」は,対象物が設計・製造 された時点で既に有している「機能」であるとも表現する ことができる。

本研究では、認識対象を生活支援ロボットが通常扱うで あろう日用品に限定し、この機能を認識することを目的と する。一般的には、このような機能は対象物の部分的な形 状領域ごとに与えられる。例えばコップに関しては、"飲み 物を蓄える"という機能は大きなくぼみ領域に付与されてお り、"(手で)つかまれる"という機能はアーチ型または棒状 の領域に付与されている。したがって本論文では,対象物 の部分領域ごとに,それがもつ「機能」の属性ラベルを適 切に割り当てる手法を提案する。

なお日用品が持っている機能は、大きなくぼみのような 特徴的な形状によって発現するものだけでなく、ナイフの 刃(属性=「切る」)のように材質によって発現するもの、 フォークの先端(属性=「刺す」)のように動作方向によっ て発現するものなどがあるが、本研究では最も一般的な、 形状由来の機能のみを扱う。以上のことから、本研究で扱 う機能は部分領域ごとに付与されるものであり、なおかつ、 形状に由来するものであるという前提を置く。

従来の研究では,機能属性を画素ごとに認識する手法が 提案されており、アドホックに設計された特徴量を学習して 得られた識別器を使用する手法⁽³⁾⁽⁴⁾, SegNet⁽⁵⁾ や, FCN⁽⁶⁾ に代表されるような Deep Learning ベースのセマンティッ クセグメンテーション手法(7)(8)が存在する。しかし、これら の手法では画素ごとに認識するために、部分領域と物体形 状を十分に考慮できていないという問題があった。そのた め昨今では、これらの手法に加えて dense CRF (Fully Connected Conditional Random Field)⁽⁹⁾ を後段に加えることに よって, 画素間のつながりを考慮して認識結果を最適化す る手法が提案されている(10)(11)。dense CRF は各画素の周辺 領域をガウスマップ的に抽出しその領域内で最適化をおこ なうため, 画素ごとの認識に比べて部分領域を考慮できる ことが期待される。しかし、従来手法(10)(11)では、最適化に おいて RGB 情報のみを用いていることから、くぼみやアー チ型など、3次元的に表現される物体形状を考慮できない という問題があった。

そこで本研究では、画素間の3次元的な位置関係を考慮 した dense CRF を用いることによって、RGB データだけで は推定できなかった面の滑らかさなどを考慮し、機能属性 認識を高精度化する。物体形状と機能の間の関係は、同一 機能を持つ領域において、従来手法で考慮されていた、類 似した色で構成されているという特徴と同時に、滑らかな 面で構成されているという特徴をもつと考えられる。例え ば、Fig.1のスプーンに関しては、"Grasp"と"Scoop"それ ぞれの領域内では同一の色特徴を持つ。また、同一の機能 を持つ領域は滑らかな面を持っていることに対し、異なる



(b) Example of functional attribute

Fig. 1. Example of functional attribute in everyday objects.

機能の境界には3次元的なエッジが生じている。これらを RGB-D データから考慮すると,滑らかな面は対象の領域 内に存在する法線情報から評価でき,類似した色は領域内 に存在する画素のRGB情報によって評価できる。本研究 では,これらの評価方法をもとに dense CRF を再設計する ことによって,機能がもつ特徴である,RGB情報と物体形 状,部分領域の3つをすべて考慮し,機能属性の認識精度 を向上させる。

以下,第2章では本研究で取り扱う機能属性について整理し,第3章では物体形状に作り込まれた機能属性と dense CRF との関係について議論する。また,第4章で本手法の流れと使用したモデル式について詳しく述べ,第5章で実験結果と考察を記載する。そして,第6章で本論文をまとめる。

2. 本研究で取り扱う機能属性

日用品に備わる機能属性は, Nguyen や Myers らによっ てさまざま定義されている (3)(4)(7)。例えば、飲み物を蓄える という機能を有する"Contain",物を叩くという機能を有す る "Pound", 物を切るという機能を有する "Cut"などがあ る。これらに共通して存在する特徴は、これらの機能を認 識するにあたって、それぞれがもつ特有の形状が主な手掛 かりとなるということである。例えば、"Contain"はくぼみ 状, "Cut"は鋭いエッジが機能を認識する際の重要な手掛か りとなる。一方で、物体形状に加えて、その他の情報が主な 手掛かりとなる場合も考えられる。例えば、"Pound"は打撃 面の形状に加えて振り下ろすという動作, "Cut"は金属やセ ラミックスという材質がそれぞれ手掛かりとなる。しかし, いずれも形状情報に付加される情報であるため、ベースと なる情報は形状にあると考えられる。このことから、本論 文では物体形状を主な手掛かりとして想起される機能につ いて取り扱う。

3. 物体形状に基づいた機能と dense CRF との関係

本手法では、画素ごとに求められる機能属性の認識結果 を事前確率として, dense CRF を用いて認識結果の最適化 をおこなう。このとき、機能と関係性の強い物体形状を表 現するために, dense CRF で使用するデータ構造は3次元 の点群データとする。特定の機能を持つ領域内に存在する 特徴について, Fig.1 に示す機能属性の例から考える。これ らの機能は、画素ごと(あるいは3次元点群の点ごと)に 付与されているものではなく、特定の領域ごとに備わって いる。そのために、ある機能が局所的な領域において孤立 する可能性は低いといえる。また、滑らかに繋がる面内で は同一の機能属性である確率は高いと考えられる。さらに、 Fig.1のコップの例に注目すると、"Contain"を持つ領域は 類似した色で表現されているように、色情報も機能属性を 分ける手がかりとなり得ると想定できる。これらの仮定を もとにデータ構造として持つ特徴を考えると、同一の機能 属性を持つ条件は以下で表現することができる。 条件1 注目点の周辺に位置する点群のほとんどが同一の



Fig. 2. Flow of proposed method.

機能属性である場合

- 条件2 注目点とその周辺に位置する点群が持つ色特徴が 類似している場合
- 条件3 注目点とその周辺に位置する点群が持つ法線の方 向が類似している場合

本手法で使用する dense CRF はこれらの条件を評価する。 具体的な評価式は,4.2 節にて述べる。

4. 提案手法の流れ

提案手法の流れを Fig.2 に示す。入力データは 3D センサ から得られる単視点の RGB-D データ(点群データ)であ る。提案手法では、点群データの全点を利用すると処理コ ストが高くなるために,前処理として複数点を1つの局所 的な領域となるように、入力データを過分割する。これに は, Supervoxels Segmentation 手法である VCCS 法⁽¹²⁾ を用 いる。この手法では、等間隔にシード点を配置し、それぞ れのシード点をベースとして色と法線、および各シード点 と近傍点との距離の情報から類似する点を統合し、局所領 域を作成する。なお、本手法においては、最初に2cmおき にシード点を分布させることによって, 局所領域の大きさ も約2cm 矩形になるように定めている。次に,各局所領域 から数点をランダムで抽出し, Random Forests を用いてそ れぞれ機能属性の尤度マップを推定する。そして、この尤 度マップを利用して,局所領域ごとの機能属性の尤度マッ プを確定する。こうして得られた局所領域ごとの尤度マッ プから, dense CRF によって全体最適化をおこなう。最終 的に得られる出力としては,局所領域ごとに一意に機能属 性が割り当てられた点群データであり、これは dense CRF を用いて最適化された尤度マップをもとに MAP 推定をお こなうことによって取得する。

〈4・1〉 形状特徴量 Random Forests を用いて学習する形状特徴量には、物体形状を表現可能な Shape Index⁽¹³⁾と Curvedness⁽¹⁴⁾を利用する。Shape Index は [0, 1] で連続するスカラー値で物体形状を表現する特徴量であり、(1)式によって定義される。

ここで, κ_1, κ_2 は主曲率を示し, $\kappa_1 \ge \kappa_2$ が成り立つ。本 研究では,特徴量記述空間においてヘッセ行列を算出する ことによって,この主曲率を求める。この特徴量は,ドー ム型やお椀型,サドル型などの形状を連続値として推定す ることが可能である。しかしながら,(1)式から読み取れ るように,2つの主曲率の差分から形状を表現しているた め、主曲率の絶対値の情報が欠落し、推定された形状の曲 がり度合いを推定することができない。そこで、この Shape Index と組み合わせて、(2) 式で定義される Curvedness を 利用する。

本手法では、この2つの特徴量をスケールの異なる5つの特徴量記述範囲から推定し、10次元の形状特徴量として利用する。

〈4・2〉 dense CRF 最適化のために用いる dense CRF は式 (3) で定義される⁽⁹⁾。

$$E(x) = \sum_{i} \theta_{i}(x_{i}) + \sum_{i < j} \theta_{i,j}(x_{i}, x_{j}) \cdots \cdots \cdots \cdots \cdots (3)$$

式中の*i*, *j* は, それぞれ注目領域と探索領域それぞれの ID を示しており, x_i は注目領域 *i* に付与した機能属性を示し ている。また, $\theta_i(x_i)$ は x_i のときの Unary Potential を表し, これは各局所領域に付随する事前確率に該当する。本手法 においては, 機能属性の事前確率として, Random Forests による識別によって得られる各機能属性の尤度を $P(x_i)$ と したとき, $\theta_i = -\log(P(x_i))$ とする。 $\theta_{i,j}(x_i, x_j)$ は Pairwise Potential を表す。これは,注目領域と比較する領域との特 徴の違いをさまざまな観点から評価する関数である。本手法 では,第2章で仮定した条件を組み込んだ(4)式を用いる。

$$\theta_{i,j}(x_i, x_j) = \mu(x_i, x_j) \bigg[w_1 \exp \left(-\frac{\| \mathbf{p}_i - \mathbf{p}_j \|}{2\sigma_p^2} \right) \\ + w_2 \exp \left(-\frac{\| \mathbf{p}_i - \mathbf{p}_j \|}{2\sigma_{pI}^2} - \frac{\| \mathbf{I}_i - \mathbf{I}_j \|}{2\sigma_I^2} \right) \\ + w_3 \exp \left(-\frac{\| \mathbf{p}_i - \mathbf{p}_j \|}{2\sigma_{pn}^2} - \frac{\| \mathbf{n}_i - \mathbf{n}_j \|}{2\sigma_n^2} \right) \bigg]$$
(4)

 $\mu(x_i, x_j)$ は $x_i \neq x_j$ のとき1となり,それ以外は0となる 関数を示す。wは重み変数を, σ は分散を示す。また,変 数p,I,nはそれぞれ対象領域の重心,代表となる色情報, 法線情報を示す。そのため、w1に係る関数は,注目領域と 比較領域との3次元的な距離を表しており,この評価関数 によって機能属性認識結果における孤立点を抑制する働き を持ち,前述の条件1に相当する。w2に係る関数は,各領 域の3次元的な距離と色情報の類似度合を評価した関数で あり,これは条件2に相当する。w3は各領域の3次元的な 距離と法線の類似度合を評価した関数であり,条件3に相 当する。

5. 実験結果と考察

機能属性認識の実験では、17 種類の日用品にそれぞれ 機能属性のラベルが付与されているデータセット(RGB-D Part Affordance Dataset⁽³⁾)を用いた。Fig. 3 にこのデータ セットの例を示す。このデータセットは、機能属性認識の 関連手法⁽³⁾⁽⁴⁾⁽¹⁾においても共通的に用いられているもので



Fig. 3. Example images of RGB-D Part Affordance Dataset⁽³⁾.



■ 0.7-0.71 ■ 0.71-0.72 ■ 0.72-0.73 ■ 0.73-0.74 ■ 0.74-0.75 ■ 0.75-0.76 ■ 0.76-0.77 ■ 0.77-0.78 ■ 0.78-0.79 ■ 0.79-0.8

あり、本研究ではこのデータセットのうち、第2章にて認 識対象と定めた形状のみに由来する Contain, Grasp, Wrap-Grasp, Scoop, Support を選択し、この5種類の属性を持つ 8種類の日用品を認識対象とした。また、学習用と識別用の データの分類は、あらかじめ使用したデータセットにて定 められた方法に従った。実験の環境は、CPU: Intel ® Xeon ® E3-1240, RAM: 32 GB でおこなった。なお、dense CRF による最適化回数は 20 回とした。

〈5・1〉 denseCRFにおける重みパラメータと認識率との関係 本節では、(4)式における重みパラメータと認識精度との関係について考察する。この関係を明らかにするために、3つの重みパラメータを[0.0,5.0]の間で0.2ずつ変化させた全26³通りで認識率を算出し、それらの認識率と各パラメータの値の関係を分析した。Fig.4にはその結果の一例として、w1=1に固定し、w2とw3、認識率の3つの関係を可視化した結果を示す。この結果より、各パラメータが認識精度に貢献していることがわかる。また、w1についても、同様の可視方法によって分析した結果より精度向上に貢献していることを確認した。なお、提案手法においては、この実験から認識率が最大となるようにw1=1.0,w2=2.0,w3=3.0と設定した。

(5・2) 機能属性認識の評価実験 Fig.5 に各手法の認識結果例を示し, Fig.6 に Confusion Matrix として表した 各手法の認識結果を示す。Table 1 には, 各手法の平均認識 率を示す。なお,本実験における認識率は, Ground truth に ラベル付けされた点群のうち,正しく機能属性が認識でき た割合を示している。また,比較手法として用いた手法は

Table 1.	Mean re	cognition	rate of	each	method
----------	---------	-----------	---------	------	--------

	w/o dense CRF	2D dense CRF	3D dense CRF (Proposed)
Mean	62.4	73.2	77.0



Fig. 5. Example of recognition result.

Fig. 4. Evaluational experiment for each weight parameter.



Fig. 6. Recognition result expressed by confusion matrix.



Fig. 7. A failure case when the shovel is recognized.

2 つあり、1 つ目は dense CRF による最適化をおこなわず に局所領域ごとの機能属性の認識結果をそのまま利用した 手法(w/o dense CRF)である。2 つ目は、従来手法(")に おいて、RGB データのみから dense CRF を用いた最適化 を模擬し、画素ごとの機能認識は提案手法と同一の処理に 統一した手法(2D dense CRF)である。なお、実験結果の 図および表においては、提案手法を 3D dense CRF と表記 した。

Fig.6から各手法を比較すると,手法(a)に比べて手法(b), (c)のほうが対角線上の値が高くなっているために,dense CRFを使用することによって機能属性の認識率が向上した ことがわかる。また,従来手法の手法(b)と提案手法の手 法(c)を比較すると,提案手法のほうが"Grasp"以外の機 能属性について認識率が高いことが確認できる。さらに, Table 1の平均認識率から確認すると,提案手法の平均認識 率は77.0%であり,これはRGBデータのみを用いたdense CRF による機能属性の最適化手法よりも 3.8%高い認識率 であった。

一方で,提案手法の中でも最も認識率が低かった機能属 性は "Support"であり,認識率は 60.4%であった。これは, Fig. 7 に示すように Shovel に備わる "Support"を "Scoop" と誤認識したことが多く見られたことに依存していると考 えられる。この誤認識は, Shovel に備わる "Support"が持 つ形状が, Fig. 5 の右上の Trowel に備わる "Scoop"が持つ 形状と類似していることが原因であると考察する。これを 解決するためには,あらかじめ対象物の物体名を認識し,ク ラス名から想定される機能を認識する処理を加えるという 方法が有効であると考えられる。

なお、本手法では機能属性認識に使用する点群を局所領 域ごとに過分割することによる処理コスト削減と、Random Forests を利用した高速な識別の組み合わせによって,副次 的効果として認識処理速度が高速であることも確認した。提 案手法による一物体あたりの平均の認識処理時間は109 ms であった。

6. まとめ

本論文では、日用品の機能属性を高精度に認識する手法 を提案した。提案手法では、Random Forests による機能属 性認識結果を事前確率として、物体形状を表現した dense CRF を用いることによって、全体最適化をおこなった。実 験の結果、提案手法の認識率は 77.0%であり、これは従来 よりも 3.8%高い結果であった。また、本手法では、入力 データを過分割することによる処理コスト削減と、Random Forests による高速な識別によって、副次的効果として処理 速度が高速であることも確認した。提案手法の処理速度は 109 ms であった。

謝 辞

この成果の一部は、国立研究開発法人新エネルギー・産業 技術総合開発機構(NEDO)からの委託業務の結果である。

文 献

- J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: "You Only Look Once:Unified, Real-Time Object Detection", In Proc. IEEE Conf. of CVPR, pp. 779-788, Las Vegas, USA (2016)
- (2) W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A.C. Berg: "SSD: Single Shot MultiBox Detector", In Proc. European Conf. on ECCV, pp.21-37, Vancouver, Canada (2016)
- (3) A. Myers, C.L. Teo, C. Fermuller, and Y. Aloimonos: "Affordance Detection of Tool Parts from Geometric Features", Proc. IEEE Conf. on ICRA, pp.1374-1381, Washington, USA (2015)
- (4) S. Akizuki, M. Iizuka, and M. Hashimoto: "Affordance"-focused Features for Generic Object Recognition", European Conf. on ECCV Workshop, Vancouver, Canada (2016)
- (5) V. Badrinarayanan, A. Kendall, and R. Cipolla: "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", Trans. on PAMI, Vol.39, No.12 (2017)
- (6) J. Long, E. Shelhamer, and T. Darrell: "Fully Convolutional Networks for Semantic Segmentation", Proc. IEEE Conf. of CVPR, pp. 3431-3440, Massachusetts, USA (2015)
- (7) A. Nguyen, D. Kanoulas, D.G. Caldwell, and N.G. Tsagarakis: "Detecting object affordances with Convolutional Neural Networks", Proc. IEEE/RSJ Conf. on IROS, pp.2765-2770, Daejeon, Korea (2016)
- (8) A. Roy and S. Todorovic: "A Multi-Scale CNN for Affordance Segmentation in RGB Images", Proc. IEEE Conf. on ECCV, pp. 186-201, Vancouver, Canada (2016)
- (9) P. Krahenbuhl and V. Koltun: "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials", Proc. Conf. on NIPS, pp.109-117, Granada, Spain (2011)
- (10) L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille: "Semantic image segmentation with deep convolutional nets and fully connected crfs", Proc. Conf. on ICLR, California, USA (2015)
- (11) A. Nguyen, D. Kanoulas, D.G. Caldwell, and N.G. Tsagarakis: "Object-Based Affordances Detection with Convolutional Neural Networks and Dense Conditional Random Fields", Proc. IEEE/RSJ Conf. on IROS, Canada (2017)
- (12) J. Papon, A. Abramov, and M. Schoeler: "Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds", Proc. IEEE Conf. on CVPR, pp.2027-2034, Oregon, USA (2013)
- (13) C. Dorai and A.K. Jain: "COSMOS-A Representation Scheme for 3D Free-Form Objects", IEEE Trans. on PAMI, Vol.19, No.10, pp.1115-1130 (1997)
- (14) J.J. Koenderink, and A.J. van. Doorn: "Surface shape and curvature scales", Trans. on IVC, Vol.10, No.8, pp.557-564 (1992)



飯 塚 正 樹 (非会員) 2017 年 4 月中京大学大学院工学研究 科機械システム工学専攻入学。博士前期課程在籍。 ロボットビジョン,3次元物体認識に興味を持つ。 2017 年度 IEEE Nagoya Section Excellent Student Award, 2017 年度画像センシングシンポジウム優 秀学術賞等受賞。日本ロボット学会,情報処理学 会各会員。



学 (正員) 1987年大阪大学大学院工学研究科 (溶接 工学専攻)博士前期課程修了。同年三菱電機(株) 入社。先端技術総合研究所等にてロボットビジョ ン,3次元物体認識、ヒューマン認識技術などの 研究開発に従事。2008年中京大学情報理工学部 機械情報工学科教授。2013年より同大学工学部教 授。2017年より工学部長。博士(工学)。1998年 度日本ロボット学会実用化技術賞, 2012/2017年

度画像センシングシンポジウム優秀学術賞,精密工学会画像応用技術 専門委員会小田原賞, 2017/2018 年 IWAIT Best Paper Award 等受賞。 電子情報通信学会,日本ロボット学会,IEEE 各会員。



秋 月 秀 一 (正員) 2016 年 9 月中京大学大学院博士後期課 程修了。日本学術振興会特別研究員 (DC2, PD) を 経て、2017年より慶應義塾大学助教(有期)、中京 大学人工知能高等研究所特任研究員。博士 (情報 科学)。3次元物体認識技術,知能ロボティクスの 研究に従事。2017年度精密工学会技術奨励賞等 受賞。電子情報通信学会等会員。