

AI時代の3次元ロボットビジョン

橋本 学
Manabu Hashimoto

中京大学 工学部
School of Engineering, Chukyo University

ABSTRACT

In this manuscript, we introduce 3D robot vision technologies which can be utilized in the fields of manufacturing, logistics and home. Versatile commercial range sensors are overviewed, and several efficient 3D local features for key-point matching are also explained. For future robot technologies, we propose a new robot sensing method which is based on “function” recognition. The robot can generate natural and appropriate motion path by using recognized results of “functions”, because that will provide us beneficial information about usage of objects to be operated.

Keywords: 3次元, 物体認識, ロボットビジョン, 機能認識, 生活支援ロボット

1. はじめに

ロボットビジョン技術は、生産、物流、家庭等のさまざまな状況下で必要な技術であり、既に普及率の高い分野もあるが、家庭における生活支援ロボットに関しては、まだ発展途上の部分も少なくない。本稿では、3次元データ、とりわけポイントクラウドデータ（点群データ）を想定しつつ、ロボットビジョンの課題を整理した上で、基本的な物体認識技術を紹介するとともに、AI時代におけるロボットビジョンの展望として、最近注力している「機能認識」に基づくロボット動作生成等の新技術を紹介する。

2. ロボットビジョンの課題

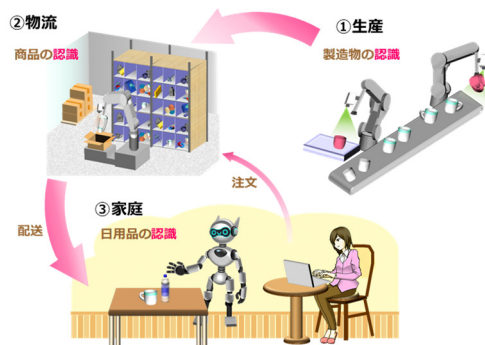


図1. 生産・物流・家庭におけるロボットビジョン

図1は、生産、物流、家庭の各場面におけるロボットビジョンの使用例である。製造物が工場で生産され（①）、家庭からECにて発注された商品として物流・配送センター（②）にて梱包、輸送され、家庭に届く。そこでは、生活支援ロボット（③）が日用品を把持し、ハンドリングする。このように、いたるところでロボットビジョンが使用される。この3つの場面において、ロボットビジョンに課された課題が異なる（図2）。



図2. さまざまな分野における物体認識の課題

たとえば、認識時における対象物の3次元形状モデルの利用可否という観点からは、生産や一部の物流分野では利用可能であるが、家庭における生活支援ロボットにおいてはモデルの利用はほとんどの場合は期待できない。実際、生産現場では取り扱う対象物は品種数も少なく、既知であることが多いので、ロボットの目の前に「何があるか」ではなく「どこにあるか（位置・姿勢認識）」が重要である。これを特定物体認識と呼んでいる。一方、生活支援ロボットが活躍する場面では、日用品などきわめて他品種の対象物に囲まれて

橋本 学

mana@isl.sist.chukyo-u.ac.jp

中京大学 工学部

〒466-8666 名古屋市昭和区八事本町 101-2

いる環境であり，そこでは，個々のモデルを使わずに対象物の名前を言い当てることも必要になってくる（一般物体認識）。

またロボットのタスクの観点からは，生産や物流では，ロボットが対象物を把持しさえすればよいことが多く，実際にはその後に把持した対象物を離すが，重要なことはいかにして把持するかである．これを，Pick and Place タスクと呼んでいる．一方，家庭内のロボットでは，例えばスプーンのような「道具」を把持したロボットは，その後，それを使ってなんらかの動作を行う（Pick and Operation タスク）．この2種類のタスクでは，ロボットビジョンに求められる位置精度が大きく異なり，当然ながら後者のほうが高い性能が求められることになる．

3. ロボットビジョンの現状

ロボットビジョンを構成する技術は，おおまかには2つに分けることができる．一つめはセンサ技術であり，3次元世界をデータとしてコンピュータに取り込むためのハードウェアである．従来はステレオカメラなど通常の2次元カラーカメラを使うものが多かったが，近年では対象物の3次元座標を直接的に得るためのレンジセンサの利用例が急速に拡大している．

二つめの技術は，言うまでもなく物体認識アルゴリズムである．こちらはセンサの発展に連動して変遷してきており，図3に示した2方式が主流である．

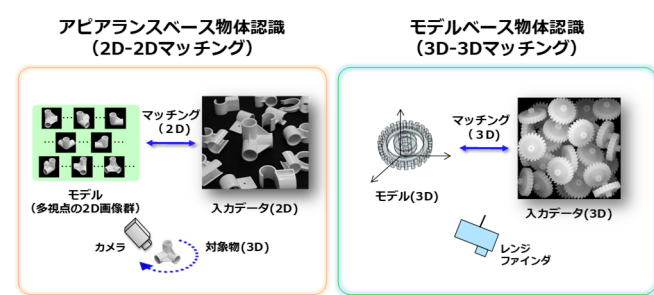


図3. 物体認識アルゴリズムの2つのタイプ

アピランススペース物体認識は，単眼カメラにより撮影されたさまざまな姿勢の対象物画像群を辞書として格納し，入力画像と辞書データを比較照合することによって対象物の姿勢を同定する手法である．近年は後述するモデルベース手法の陰に隠れがちではあるが，小型のカメラのみで実現できることから，ロボットシステムにとっては，ハンドアイシステムを構成しやすい，高解像度が得やすいなど，実用面での利点は捨てがたく，まだなお研究が継続している技術である．

モデルベース手法，特にキーポイントマッチング^[1]

と呼ばれる手法は，ここ数年の間に急速に進歩している．認識対象物のモデルおよび入力シーンを，いずれも3次元座標群（点群データ）として記述した上で，特徴点（キーポイント）と呼ばれる限定された個数の点のみを抽出し，各点に付与された「特徴量」情報を手がかりにして，モデルと入力シーンとの点ごとの対応付けを行う手法である．ここでは，特徴量の設計が何より重要であり，SHOT 法，PFH 法など，さまざまな手法が提案されている．表1に主要な3次元特徴量をまとめる．なお，下線を施した特徴量は，フリーで利用できる汎用3次元画像処理ライブラリ PCL (Point Cloud Library) に登録されたものである．

表1. 主な3次元特徴量

特徴量の記述方法			
(A) キーポイント周りの情報記述		(B) 複数点間の関係記述	
座標データ利用	法線ベクトル利用	その他の情報を利用	2点間 or 3点間
SHOT, PFH, CSHOT, B-SHOT, HONV		LSP, CCDoN, <u>NARF</u> , <u>MHOG</u> , <u>RoPS</u>	PPF, ER, VC-PPF, MPPF, PPF B2B or S2B or L2L, VPM
<u>SI</u> , <u>SSI</u> , <u>DAI</u> , <u>3DSC</u> , <u>SDLSD</u> , <u>USC</u>	<u>FPFH</u> , <u>DoN</u>		

PFH : Point Feature Histogram

SHOT : Signature of Histograms of Orientations

CSHOT : Color SHOT

B-SHOT: Binary SHOT

HONV : Histogram of Oriented Normal Vector

SI : Spin Image

SSI : Spherical Spin Image

DAI : Depth Aspect Image

3DSC : 3D Shape Contexts

USC : Unique Shape Context

SDLSD : Scale-Dependent Local Shape Descriptor

FPFH : Fast PFH

DoN : Difference of Normals

LSP : Local Surface Patches

CCDoN : Combination of Curvatures and Difference of Normals

NARF : Normal Aligned Radial Feature

MHOG : Mesh HOG

RoPS : Rotational Projection Statistics

PPF : Point Pair Feature

ER : Efficient Ransac

VC-PPF : Visibility Context PPF

MPPF : Multimodal PPF

PPF B2B or S2B or L2L : PPF Boundary-to-Boundary or Surface to Boundary or Line to Line

VPM : Vector Pair Matching

※下線はPCLに登録されている手法

近年では，このような物体認識を目的とした3次元データ処理アルゴリズムの研究に加えて，認識した対象物を，ロボットがどのように操作するかを考慮した研究に注目が集まっている．物体の品種名や位置・姿勢がわかっただけではロボットは動作できず，対象物のどこを把持するか，そのときのハンドやアームのアプローチ姿勢はどうあるべきか，また実際にそのようなアプローチはそのロボットにとって実現可能かどうか（ロボットの身体性制約）など，「認識」と「ロボット動作」とを円滑につなぐための技術は多い．

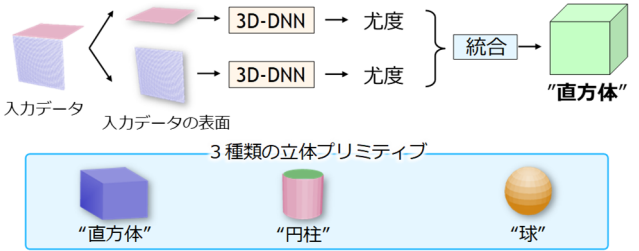


図4. 対象物のプリミティブ近似^[2]

図4に示したのは，そのようなロボット動作パラメ

ータを自動的に生成するための技術の一例である^[2]。この技術は、センサによって得られた点群データをもとに、一般的な対象物を、「直方体」「円柱」「球」の3種類の基本形状（プリミティブ形状）に近似することによって、個々の対象物の具体的な形状モデルを利用することなく、初めて目にした対象物に対しても、適切な把持位置および把持パラメータ（ハンドのアプローチ方向など）を生成するものである。我々はこの技術を、物流分野の実用技術を競う国際ロボット大会（Amazon Robotics Challenge 2017）に投入し、好成績を収めている。

4. AI時代のロボットビジョンの展望

家庭で人間の指示を受けて人間を支援するロボットについては、たとえば「ジュースを持ってきて」のような人間による極めてシンプルな要求を、複雑なロボットマニピュレーションタスクとして実際に実行することが必要である。これを全自動で実行するためには、要求を受けてから実行するまでの手順を生成する必要があるが、生活シーンという日常環境を考えると、ロボットビジョンの観点からは、3次元形状モデル等の個々の対象物に関する事前知識をほとんど使わずにおこなうことが求められるという点において、特段に難度の高い課題となっている。

上記の例では、人間がロボットに対して「コップにジュースを入れて持ってきて」という指示をした場合、図5のように、ロボットは主として、ボトルを持つ、ジュースを注ぐ、コップを持つという3つのサブタスクを実行する。これらはさらに細かく分割できる。たとえば、「ジュースを注ぐ」場合は、コップを見つける、注ぐべき位置を判断する、ボトルを傾けるという、少なくとも3つの動作を実行することになる。



図5. ロボットタスクを構成する動作

このうち、コップを見つけるような識別タスクは、いわゆるディープラーニングのような画像ベースの識別器の急激な発展により、かなり容易に実現できるようになってきた。またボトルを傾けるという基本的ロ

ット動作生成については、作用点が決まっていればさほど難易度は高くない。このことから、これらのサブタスクを実現するためには「持つべき位置」や「注ぐべき位置」を判断すること、すなわち対象物の「機能」認識が最重要であると考えられる。

一般に日用品とは、何らかの特定の機能を有している。その期待されている機能を実現するために、特に局所的な形状が作り込まれていることが少なくない^[3]。図6に、日用品が持つ機能の例を示す。

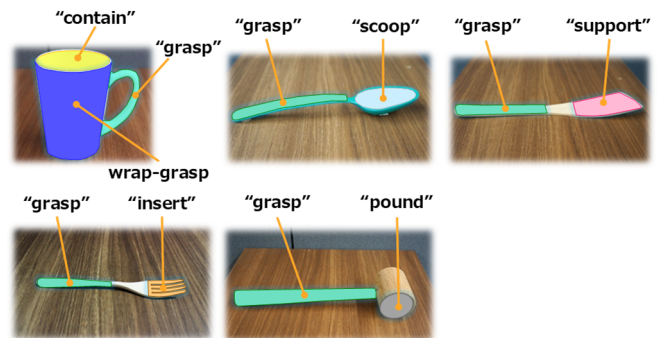


図6. 日用品が持つさまざまな「機能」の例

例えば、コップには、本来の使用目的である水（液体）を蓄えるということを実現するために、大きくくぼみ形状の部位には「contain」という機能が備わっていると考えられる。同時に、人間が手でつまむことを想定して取って部分には「grasp」という機能が、また胴体部分については人間が手のひらで掴むことができることから、「wrap-grasp」という機能が備わっていると考えられることができる。

我々は、3次元形状に起因して付与されている機能を認識する手法を開発した^[4]。この手法は、図7に示されるような機械学習ベースのものである。

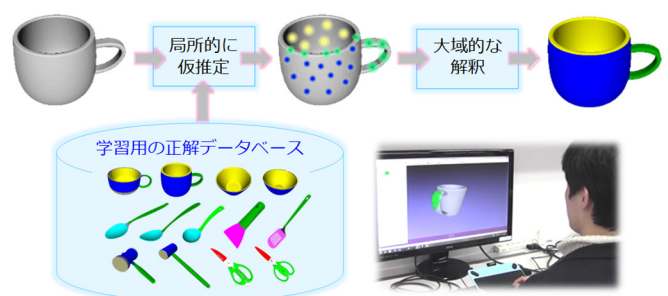


図7. 機能認識アルゴリズムの概要

まず、学習用の正解データベースを作成する。3次元形状モデルに対し、事前に部位ごとに機能ラベルを付与したデータを準備しておく。一方、レンジファインダから得られた未知対象物のポイントクラウドデータに対し、このデータベースを適用して対象物上の局所

領域ごとに機能 ID に関する仮ラベルを推定する．具体的には，ランダムフォレストやディープラーニングを利用している．次に，この仮ラベル群を統合し，最終的な機能ラベリングを実現する．

ここで，初期段階における仮ラベル付与の単位としての局所領域の大きさが重要となる．提案手法では，図 8 に示した程度の大きさを想定している．最終的なラベル領域より十分小さいものの，ポイントクラウドデータを数十点以上を含む程度の領域が望ましい．この局所領域ごとに仮ラベルを付与し，これを統合する過程において，たとえば，隣接領域のラベルは同一であるべきなどの制約を満たすように一部の尤度の低いラベルを修正し，最終的なラベル結果を得る．この方法としては，dense CRF を利用している．



図 8. 仮ラベル統合による機能ラベリングの高信頼化

この手法による機能認識結果を図 9 に示す．コップやへら（スパチュラ）等に対して，提案手法を適用した結果，Step1 では局所領域単位の仮ラベルに付与ミスが散見されるが，Step2 を経てそれらが低減し，真値に近いラベリング結果が得られていることがわかる．

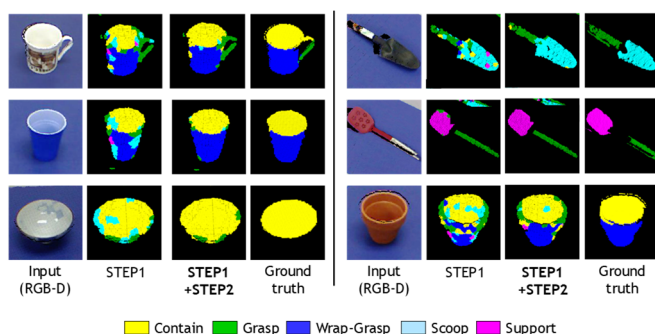


図 9. 機能認識の実験結果

なお，提案手法による処理時間としては，一般的なデスクトップ PC による CPU 処理により，約 15fps の速度が得られており，これはロボットタスク実行の観点からは，ほぼリアルタイム処理と言ってよい程度の性能である．

機能認識技術の活用事例として，把持パラメータ推定手法を紹介する^[5]．図 10 に示したように，レンジファインダから得られた RGB-D データをもとに，まず機能認識をおこなうと同時に，D データを用いた把持パラメータの候補群を得る．この段階では，把持可能な候補が複数得られる．この候補群のなかから最適なパラメータを決定する際に，対象物の局所形状に対して決定済の機能ラベルを利用する．たとえば，“grasp”ラベルが付与された部位に設定された把持位置は正しいが，“contain”ラベルが付与された部位に設定された把持位置は正しくない（好ましくない）と判断される．

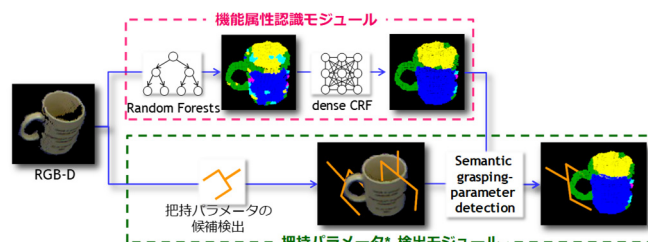


図 10. 機能認識を利用した把持位置推定

5. おわりに

本稿では，いわゆる点群データ処理に基づくロボットビジョン技術について，基本課題を明らかにした上で，キーポイントマッチングをはじめとする諸技術を概観し，来たるべき AI 時代の高度ロボットシステムのための新技術として「機能認識」とそれを応用したロボットビジョンの研究事例を紹介した．

文献

- [1] 橋本学ほか，物体認識のための 3 次元特徴量の研究動向，電気学会論文誌 C，Vol. 136，No. 8，pp. 1038-1046，2016/8.
- [2] T. Torii, M. Hashimoto, Model-less Estimation Method for Robot Grasping Parameters Using 3D Shape Primitive Approximation, IEEE CASE2018, pp. 580-585, 2018.
- [3] 山野辺夏樹ほか，ロボットマニピュレーション研究におけるアフォーダンスに関するレビュー，日本ロボット学会誌，Vol. 36，No. 5，pp. 327-337，2018.
- [4] 飯塚正樹，秋月秀一，橋本学，物体形状を考慮した denseCRF による機能属性認識の高精度化，電気学会論文誌 C，Vol. 138，No. 9，pp. 1088-1093，2018.
- [5] M. Iizuka, M. Hashimoto, Detection of Semantic Grasping-Parameter using Part-Affordance Recognition, 19th International Conference on REM, pp. 136-140, DOI:10.1109/REM.2018.8421780, 2018.